A korarchaeal genome reveals insights into the evolution of the Archaea

James G. Elkins*[†], Mircea Podar[‡], David E. Graham[§], Kira S. Makarova[¶], Yuri Wolf[¶], Lennart Randau[||], Brian P. Hedlund**, Céline Brochier-Armanet[††], Victor Kunin[‡‡], Iain Anderson[‡‡], Alla Lapidus[‡‡], Eugene Goltsman[‡‡], Kerrie Barry[‡‡], Eugene V. Koonin[¶], Phil Hugenholtz[‡‡], Nikos Kyrpides[‡‡], Gerhard Wanner[§§], Paul Richardson[‡‡], Martin Keller[‡], and Karl O. Stetter*[¶¶||||]

*Lehrstuhl für Mikrobiologie und Archaeenzentrum, Universität Regensburg, D-93053, Regensburg, Germany; [‡]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831; [§]Department of Chemistry and Biochemistry, The University of Texas at Austin, Austin, TX 78712; [¶]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894; [||]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520; **School of Life Sciences, University of Nevada, Las Vegas, NV, 89154; **Université de Provence - Aix-Marseille I, Laboratoire de chimie bactérienne, UPR CNRS 9043, Marseille, France; [‡‡]DOE Joint Genome Institute, Walnut Creek, CA 94598; [§§]Institute of Botany, LM University of Munich, D-80638, Munich, Germany; [¶¶]Institute of Geophysics and Planetary Science, University of California, Los Angeles, CA 90095

[†]Present address: Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831

[||||]To whom correspondence should be addressed:

Karl O. Stetter
Universität Regensburg
Lehrstuhl für Mikrobiologie
Universitätstraße 31
93343 Regensburg
Phone: +49-941 943-1821, -3161
Fax: +49-941 943-3243
E-Mail: karl.stetter@biologie.uni-regensburg.de

Manuscript info: 13 pages, 2 figures, 2 tables.

Abbreviations: SSU, small subunit; LSU, large subunit; arCOGs, archaeal clusters of orthologous groups; RNAP, RNA polymerase; EF, elongation factor; HGT, horizontal gene transfer.

The genome sequence describe here has been deposited in GenBank under accession no. CP000968.

Abstract:

The candidate division *Korarchaeota* comprises a group of uncultivated microorganisms which, by their small subunit rRNA phylogeny, may have diverged early from the major archaeal phyla *Crenarchaeota* and *Euryarchaeota*. Here we report the initial characterization of a member of the *Korarchaeota* with the proposed name, "*Candidatus* Korarchaeum cryptofilum" which exhibits an ultra-thin, filamentous morphology. To investigate possible ancestral relationships between deep-branching *Korarchaeota* and other phyla, we used whole-genome shotgun (WGS) sequencing to construct a complete, composite korarchaeal genome from enriched cells. The genome was assembled into a single contig 1.59 Mb in length with a G+C content of 49%. Of the 1617 predicted protein-coding genes, 1382 (85%) could be assigned to a revised set of archaeal COGs. The predicted gene functions suggest that the organism relies on a simple mode of peptide fermentation for carbon and energy and lacks the ability to synthesize *de novo* purines, coenzyme A, and several other cofactors. Phylogenetic analyses based on conserved single genes and concatenated protein sequences positioned the korarchaeote as a deep archaeal lineage with an apparent affinity to the *Crenarchaeota*. However, the predicted gene content revealed that several conserved cellular systems such as cell division, DNA replication, and tRNA maturation resemble the counterparts in the *Euryarchaeota*. In light of the known composition of archaeal genomes, the *Korarchaeota* might have retained a set of cellular features that represents the ancestral archaeal form.

Introduction:

Two established phyla, the *Crenarchaeota* and *Euryarchaeota*, divide the archaeal domain based on fundamental differences in translation, transcription and replication (1). Yet hydrothermal environments have yielded small subunit (SSU) rRNA gene sequences that form deep-branching phylogenetic lineages, which potentially lie outside of these major groups. These uncultivated organisms include members of the *Korarchaeota* (2-4), the Ancient Archaeal Group (5), and the Marine Hydrothermal Vent Group (5, 6). The Nanoarchaeota have also been suggested to hold a basal phylogenetic position (7) but this placement has been debated (8). The *Korarchaeota* comprise the largest group of deep-branching, unclassified archaea and have been detected in several geographically isolated terrestrial and marine thermal environments (2, 4, 5, 9-17).

To gain new insights into the *Korarchaeota*, we revisited the original site where Barns et al. (2) collected the first korarchaeal environmental SSU rDNA sequences (pJP78 and pJP27) from Obsidian Pool, Yellowstone National Park, Wyoming, USA. Continuous enrichment cultures were established at 85^{o} C using a dilute organic medium and sediment samples from Obsidian Pool as an inoculum. The cultivation system supported the stable growth of a mixed community of hyperthermophilic bacteria and archaea including an organism with a SSU rDNA sequence displaying 99% identity to pJP27. The organism was identified as an ultra-thin filament between 0.16-0.18 μm in diameter and variable in length. Whole-genome shotgun (WGS) sequencing was used to assemble an intact composite genome from purified cells originating from the enrichment culture. The complete genome sequence of "*Candidatus* (*Ca.*) Korarchaeum cryptofilum" provides the first look into the biology of these deeply-branching archaea and their evolutionary relationships with *Crenarchaeota* and *Euryarchaeota*.

Results:

Cultivation and *In-situ* Identification. An enrichment culture was inoculated with sediment and hot spring samples taken from Obsidian Pool, YNP. The enrichment was maintained under strictly anaerobic conditions at 85^{o} C, pH 6.5 and continuously fed a dilute organic medium. A stable community of hyperthermophilic archaea and bacteria with a total cell density of approx. $1.0x10^{8}$ cells/ml was supported for nearly 4 years. Sequences from SSU rDNA clone libraries derived from the enrichment were closely related to other known isolates or environmental sequences from Obsidian Pool [see supporting information (SI) *Text* and SI Fig. 3]. The *Korarchaeota*, were represented by the SSU rDNA clone pOPF_08 which is 99% identical to pJP27 from Obsidian Pool, YNP (2), and pAB5 from Calcite Springs, YNP (9). FISH analysis allowed the positive identification of cells with the pOPF_08 SSU rRNA sequence. Cells hybridizing to Cy3-labeled, *Korarchaeota*-specific probes, KR515R and KR565R, were ultra-thin filaments less than 0.2 μm in diameter with an average length of 15 μm, although cells were observed with lengths up to 100 μm (Fig. 1*A*, SI Fig. 4, SI *Text*).

Cell Preparation and Genome Sequencing. It was observed that filamentous cells hybridizing to probes KR515R/KR565R remained intact in the presence of high concentrations of sodium dodecyl sulphate (SDS; up to 1%) in the hybridization buffer. This feature allowed highly enriched cell preparations to be made by exposing the Obsidian Pool enrichment culture to 0.2% (w/v) SDS (without cell fixation) followed by several washing steps and filtration through 0.45 μm syringe filters. PCR amplified, SSU rDNA sequences from SDS-treated, filtered cell preparations showed that over 99% of the clones sequenced (n=180) were identical to the SSU rDNA sequence of pOPF_08 (See SI Fig. 5). Phase contrast (Fig. 1*B*) and electron microscopy (Fig. 1*C*) showed the samples to be highly enriched for ultra-thin filamentous cells with a diameter of 0.16-0.18 μm. DNA clone libraries were constructed from both SDS and non-SDS (libraries BHXI and BFPP respectively) treated enrichment culture filtrates. A total of 23,000 and 11,520 quality sequencing reads from libraries BHXI and BFPP respectively were binned based on %GC content and read depth. Overlapping fosmid sequences containing the pOPF_08 SSU rRNA gene (SI Fig. 6) were used to guide the WGS assembly. Five large scaffolds with a read depth of 8.4-9.9X were closed by PCR (further details are provided in SI *Text*). Single nucleotide polymorphisms occur at a rate of approx. 0.2% across the genome.

General Features. The complete genome consists of a circular chromosome 1,590,757 bp in length with an average G+C content of 49% (Table 1). A single operon was identified that contains genes for the SSU and LSU rRNAs. Forty five tRNAs were identified using tRNAscan-SE (18). A total of 1617 protein coding genes were predicted with an average size of 870 bp. Of the predicted protein coding genes, 72.4% included AUG; 17.6%, UUG; and 10% had GUG for start codons. The arCOG (see below) and COG analysis, combined with additional database searches allowed the assignment of a specific biological function to 998 (62%) predicted proteins; for another 246 proteins (15%), biochemical activity but not biological function could be predicted, and for the remaining 373 (23%) proteins, no functional prediction was possible, although many of these are conserved in some other archaea and/or bacteria.

A**rchaeal COGs.** The predicted proteins were assigned to archaeal COGs [(arCOGs) (19) (see SI *Text*, SI Table 3]. Of the 1617 annotated proteins, 1382 (85%) were found to belong to the

arCOGs, a coverage that is slightly lower than the mean coverage of 88% for other archaea and much greater than the lowest coverage obtained for *Nanoarchaeum equitans* (72%) and *Cenarchaeum symbiosum* (58%). When the gene complement was compared to the strictly defined core gene sets for the *Euryarchaeota* and *Crenarchaeota* (i.e., genes that are represented in all sequenced genomes from the respective division, with the possible exception for *C. symbiosum* in the case of the *Crenarchaeota*, but are missing in at least some organisms of the other division), a strong affinity with the *Crenarchaeota* was readily apparent. Specifically, "*Ca*. K. cryptofilum" possesses 169 of the 201 genes from the crenarchaeal core (84%) but only 33 of the 52 genes from the euryarchaeal core (63%). When the core gene sets were defined more liberally, i.e., as genes that are present in more than two thirds of the genomes from one division and absent in the other division, the korarchaeote actually shared more genes with the *Euryarchaeota* than with *Crenarchaeota* (Table 2, SI Table 4). Seven proteins had readily identifiable bacterial but not archaeal orthologs, as determined by assigning proteins to bacterial COGs (20) (SI Table 5). Conceivably, the respective genes were captured via independent horizontal gene transfer (HGT) events from various bacteria. By contrast, there were no proteins specifically shared with eukaryotes, to the exclusion of other archaea. The organism lacks only 5 genes that are represented in all sequenced archaeal genomes, namely, diphthamide synthase subunit DPH2, diphthamide biosynthesis methyltransferase, predicted ATPase of PP-loop superfamily; predicted Zn-ribbon RNA-binding protein, and small-conductance mechanosensitive channel.

Energy Metabolism. The predicted gene set suggests that "*Ca*. K. cryptofilum" grows heterotrophically, using a variety of peptide and amino acid degradation pathways. At least four ABC-type oligopeptide transporters and an OPT-type symporter could import short peptides, which more than a dozen peptidases could hydrolyze into amino acids. As in *Pyrococcus* spp., pyridoxal 5′-phosphate dependent aminotransferases can convert amino acids to 2-oxoacids, while scavenging amines with α-keto-glutarate to form glutamate. Four ferredoxin-dependent oxidoreductases (specific for indolepyruvate, pyruvate, 2-oxoglutarate or 2-oxoisovalerate) could oxidize and decarboxylate the 2-oxoacids, producing acyl-CoA molecules. Four acyl-CoA synthetases can convert this thioester bond energy into phosphoanhydride equivalents. Six aldehyde:ferredoxin oxidoreductase metalloenzymes could oxidize aldehydes derived from these amino acids. Pyruvate could be degraded by this pathway or by a homolog of pyruvate formate lyase. The only terminal reduction reaction predicted from the genome sequence is hydrogen production, apparently catalyzed by two soluble [NiFe]-hydrogenases. An archaeal-type proton-transporting ATP synthase would convert proton motive force produced by anaerobic respiration into ATP. However, in contrast to the system proposed for *Pyrococcus furiosus*, "*Ca*. K. cryptofilum" lacks a membrane-bound proton-translocating hydrogenase (21). Therefore proton translocation must occur through the NADH:quinone oxidoreductase complex or a novel system that might involve homologs of the methanogen *hdrABC*-type heterodisulfide reductase complex. A ferredoxin:NADP oxidoreductase, three flavin reductases, and two electron transfer flavoproteins could mediate electron transfer to the respiratory chain. The korarchaeote also encodes a homolog of a single subunit [Ni-Fe] carbon monoxide dehydrogenase and its accessory proteins in a cluster of methanogen-like genes. Although the physiological role of these proteins in methanogens is unknown, they might confer the ability to oxidize CO produced under anaerobic conditions (22). There is no cytochrome c and no evidence of the dissimilatory reduction of sulfur, sulfite, sulfate, nitrate, nitrite, iron, formate, or oxygen. An abundance of

iron-sulfur proteins, free radical initiating enzymes and the lack of oxidases suggest a strictly anaerobic lifestyle.

Central Metabolism. A partial citric acid cycle is present which includes 2-oxoglutarate: ferredoxin oxidoreductase, succinyl-CoA ligase, succinate dehydrogenase, fumarase, malate dehydrogenase, aconitase, and malic enzyme. These enzymes could be used either for the degradation or for the biosynthesis of glutamate. The organism also encodes the components of a serine hydroxymethyltransferase and glycine cleavage system. One-carbon units from this pterin-dependent pathway are used to produce methionine from homocysteine. The genome encodes few carbohydrate transporters and no hexokinase, although it has a complete pathway for glycolysis from glucose 6-phosphate or for gluconeogenesis. There are no enzymes for the classical or modified Entner-Doudoroff pathways that are found in many *Crenarchaeota*. The organism does have a modified ribulose monophosphate pathway to produce ribose 5-phosphate (23), and a standard pyrimidine biosynthetic pathway. However, it lacks genes for purine nucleotide biosynthesis. Finally, an extensive set of UDP-sugar biosynthesis proteins and glycosyltransferases suggests the presence of glycosylated proteins and lipids. Although "*Ca*. K. cryptofilum" appears to be a proficient peptide degrader, it has an extensive set of amino acid biosynthesis enzymes (see SI *Text*). However, there are many genes missing for coenzyme biosynthesis that are conserved in most of the other *Archaea*. For coenzyme A biosynthesis, it lacks the bifunctional phosphopantothenoylcysteine synthetase/decarboxylase that is found in all other sequenced archaeal genomes except for *Nanoarchaeum equitans* and *Thermofilum pendens* (24). In addition, pathways for riboflavin, pterin, lipoate, porphyrin, and quinone biosynthesis are incomplete.

DNA Replication and Cell Cycle. For initiating chromosome replication, two distinct *orc1/cdc6* homologues and a single minichromosome maintenance protein (MCM) complex are present along with genes encoding single-stranded binding protein (RPA) and primase (PriSL). The genome encodes multiple DNA-dependent, DNA polymerases including 2 family B type enzymes and both the large and small subunits of a euryarchaeal type II polymerase. Genes for the sliding clamp (PCNA), PriSL, and a *gins15* ortholog (25), are clustered with genes for the large subunit of the type II polymerase. A simplified clamp loader complex encodes the large and small subunits of replication factor C. Predicted chromatin-associated proteins include Alba and two H3-H4 histones. Like all known hyperthermophiles, reverse gyrase is present.

"*Ca*. K. cryptofilum" possesses several genes related to the ParA/MinD family of ATPases involved in chromosome partitioning and SMC-like proteins involved in chromosome segregation. The gene for this ATPase is part of a predicted operon that also includes genes for an FtsK-like ATPase (HerA) and two nucleases, proteins that are thought to comprise the basic machinery for DNA-pumping (26). The organism appears to employ the euryarchaeal mechanism for cell division as indicated by the presence of 7 genes encoding cell division GTPases (FtsZ; SI Fig. 7) (27). One of the *ftsZ* genes is included in a conserved euryarchaeal gene cluster containing *secE*, *nusG*, and several ribosomal protein genes (28, 29). In addition, 5 paralogous *ftsZ* genes are present in a 7-gene cluster that also includes a putative adapter protein (30).

Transcription and Translation. "*Ca*. K. cryptofilum" possesses a full complement of archaeal DNA dependent RNA polymerase (RNAP) subunits. The *rpoA* and *rpoB* genes encoding the

largest subunits of the RNAP are intact. In addition to the typical archaeal RNAP subunits, a coding region of 110 amino acids was identified with limited sequence similarity to the RPB8 of the eukaryotic RNAP. Subsequent in-depth analysis has shown that an ortholog of RPB8, previously thought to be missing in archaea, is also encoded by all sequenced genomes of hyperthermophilic *Crenarchaeota* (31). The RPB8 ortholog resides in a putative operon with the eukaryotic-like transcription factor, TFIIIB. To initiate basal transcription, archaeal homologues for TATA-binding protein, transcription factor B (TFB), and transcription factor E (TFE) are present. Transcriptional regulators are of the bacterial/archaeal type, with the XRE, TrmB, ArsR, PadR-like, CopG, Lrp/AsnC, and MarA families represented in the genome.

The rRNA operon contains a SSU (16S) as well as a LSU (23S) which harbours an intron-encoded LAGLIDADG type homing endonuclease similar to crenarchaeal homologues (32). A total of 33 LSU ribosomal proteins (r-proteins) and 27 SSU r-proteins are present. Notably, r-proteins S30e, S25e, S26e, and L13e that are conserved in the *Crenarchaeota* but are absent in *Euryarchaeota* (33) were identified. In contrast, large subunit r-proteins L20a, L29, and L35ae are missing from the genome.

The tRNA set consists of one initiator tRNA and 45 non-redundant elongator tRNAs. An unusual $tRNA^{Ile}$ with an UAU anticodon is predicted to decode the ATA codon instead of a modified CAU commonly found in archaea (with the exception of *Nanoarchaeum equitans*) (34). Both selenocysteine and pyrrolysine-specific tRNAs are absent. Four tRNA genes contain an intron located one base downstream of the anticodon and one tRNA gene ($tRNA^{Ser}$ CGA) contains an intron in the D-loop. The structural splicing motifs found at all 5 exon-intron junctions and the corresponding homomeric splicing endonuclease appear to reflect the conserved splicing mechanism found in *Euryarchaeota* (35). Also similarly to some *Euryarchaeota*, the universal G-1 residue found at the 5'terminus of $tRNA^{His}$ is not encoded but is predicted to be added posttranscriptionally by a guanylyltransferase. The genome encodes archaeal aminoacyl-tRNA synthetases for all of the amino acids except for glutaminyl-tRNA formation which is mediated via the tRNA-dependent transamidation pathway using the GatD and GatE proteins (36). The LysRS is of the class I type and a homodimeric GlyRS is present. The SerRS is the common type rather than the rarer version found in some methanogens (37). On the other hand, ThrRS appears to be a bacterial type and was likely acquired through a HGT event.

Phylogeny and Evolutionary Genomics. We performed a comprehensive phylogenetic analysis based on combined large and small rRNA subunits, conserved single-gene markers, and conserved concatenated proteins. Collectively, these results demonstrate that "*Ca*. K. cryptofilum" represents a deeply diverged archaeal lineage with affinity to the *Crenarchaeota*. Combined SSU+LSU rRNA subunit trees supported a deep crenarchaeal position (Fig. 2*A*). Likewise, the maximum likelihood based phylogeny of elongation factor 2 (EF2) homologues from archaeal genomes or environmental fosmid sequences corresponded with the rRNA tree (Fig. 2*B*). Phylogenetic analysis of 33 concatenated r-proteins and 3 large RNAP subunits clustered the korarchaeote with *C. symbiosum* in a deep branch joining the hyperthermophilic *Crenarchaeota* (Fig. 2*C*). However, this grouping could be a long branch attraction artifact, and a statistical test showed that a basal position of "*Ca*. K. cryptofilum" identical to that in Fig. 2*A* and 2*B* could not be ruled out. See SI *Text* for separate RNAP subunit and r-proteins phylogenies with compatibility testing.

Discussion:

Capturing a Korarchaeal Genome. Critical improvements in the cultivation and *in-situ* identification were necessary to resolve a complete korarchaeal genome. The ultra-thin, filamentous organisms hybridizing to *Korarchaeota*-specific probes displayed a thinner and generally longer morphology than what has been previously described for pJP27-type korarchaeote (38). It is not known if the morphological discrepancies are due to differences in the enrichment conditions, hence growth rate, or if variation in cell shape occurs among closely related species. Nevertheless, SDS concentrations that are generally 5 to 50 fold higher than those that are typically required for FISH analyses of hyperthermophilic archaea (39) were necessary for optimal probe penetration. The structural integrity of "*Ca*. K. cryptofilum" in the presence of surfactants is likely attributed to the densely packed S-layer revealed through EM studies (Fig. 1*D*). Exploiting this feature allowed the filamentous cells to be sufficiently purified for WGS sequencing and assembly into a single contiguous chromosome. The proposed name for the organism reflects its elusive nature and reputation for hiding within its community and also for its ultra-thin, filamentous morphology. The proposed genus, *Korarchaeum* gen. nov., stems from the originally proposed phylum designation by Barns *et al.* (3), which is derived from the Greek noun, *koros* or *kore* meaning "young man" or "young woman" respectively; and the Greek adjective *archaios* for "ancient". The proposed species name *cryptofilum* sp. nov., is derived from the Greek adjective, *kryptos,* meaning "hidden" and the Latin noun *filum*, "a thread".

Metabolism. Determining the growth requirements in detail for "*Ca*. K. cryptofilum" was not possible since the organism could only be propagated in a complex enrichment culture. Isolation attempts using Gelrite plates, dilution series, or optical tweezers were unsuccessful. However, the major aspects of the metabolism could be reconstructed from the predicted set of protein-coding genes which suggest an obligately anaerobic, heterotrophic lifestyle with peptides serving as the principal carbon and energy source. In agreement with the predicted metabolism, the enrichment culture was supplied with peptone and traces of yeast extract as the primary carbon and energy source under strictly anaerobic conditions. Anaerobic peptide utilization is a common metabolic strategy among hyperthermophilic *Crenarchaeota* and *Euryarchaeota* (40) and has been characterized in detail in the model organism *Pyrococcus furiosus* (41-44). However, "*Ca*. K. cryptofilum" apparently differs from other known hyperthermophiles in lacking the ability to utilize exogenous electron acceptors such as oxygen, nitrate, iron, or sulfur (45). Protons appear to be the primary acceptor for ferrodoxin-shuttled electrons. To avert possible growth inhibition, removal of molecular hydrogen by flushing with N_2/CO_2 gas and also by possible hydrogen consuming members of the enrichment community such as *Archaeoglobus* and *Thermodesulfobacterium* spp. might have improved cell growth. The organism also appears to lack complete pathways for the *de novo* synthesis of several cofactors which may prevent growth in axenic cultures. These coenzymes must be scavenged from the environment or the organism has evolved alternative modes for producing them. Microbial communities forming high-density mats composed of filamentous cells have yielded the highest number of amplified korarchaeal SSU rDNAs (9-11). Some of the essential nutrients for the growth might be supplied *in-situ* by other mat-forming organisms.

Evolutionary Considerations. Independent phylogenetic analyses based on combined SSU+LSU rRNA, elongation factor 2 (EF-G/EF-2), and concatenated r-proteins + RNAP subunit sequences are compatible with the notion of the *Korarchaeota* being a deeply branching lineage with affinity to the *Crenarchaeota* (Fig. 2). This genome-based assessment corroborates a previous phylogenetic analysis based on a robust set of archaeal environmental SSU rDNA sequences (46). The apparent relationship between the *Korarchaeota* and a member of the marine group 1 *Crenarchaeota* suggested by the phylogeny of concatenated r-protein + RNAP subunits (Fig. 2*C*) is of potential interest. Based on comprehensive phylogenetic analyses and gene content comparisons, the mesophilic *Crenarchaeota* have recently been proposed to form a separate major phylum within the Archaea (47). The apparent affinity between *C. symbiosum* and "*Ca*. K. cryptofilum" presented in our analysis remains to be validated since whole-genome phylogenetic reconstructions are based on a limited number of available archaeal genomes.

The genome revealed a pattern of orthologs that suggests an early divergence within the archaeal domain. The complement of information processing and cell cycle components appears to be a hybrid, with proteins composing the ribosome and RNAP shared, primarily, with *Crenarchaeota*, whereas functions involving tRNA maturation, DNA replication/repair, and cell division being more typical of the *Euryarchaeota* (Table 2). This complex pattern could have resulted from a combination of vertically inherited traits from ancestral organisms supplemented by HGT events. Recent genome analyses have shown that genes believed to be exclusive to the *Euryarchaeota* are also present in some crenarchaeotes. For example, a type-II DNA polymerase and a divergent *ftsZ* homologue are present in *Cenarchaeum symbiosum* (48), and histones are also found in mesophilic and some hyperthermophilic *Crenarchaeota* (49, 50). It remains to be determined whether these genes were vertically inherited from a common archaeal ancestor or were acquired horizontally from members of the *Euryarchaeota* or Bacteria (51).The euryarchaeal type features found in "*Ca*. K. cryptofilum" are generally more similar to those found in thermophilic and hyperthermophilic *Euryarchaeota* (SI Table 6). The presence of several mobile elements in the genome certainly suggests that the gene content may have been influenced by HGT (SI *Text*). Sequencing additional archaeal genomes will aid in determining whether the amalgam of cren-and euryarchaeal characteristics present in the korarchaeal genome represents is an ancient feature or resulted from a combination of HGT and gene loss events. More than a decade after the *Korarchaeota* were introduced based on rDNA sequences (2, 3), identifying "*Ca*. K. cryptofilum" and sequencing its genome has provided a new perspective into the biological diversity of these elusive organisms and the genomic complexity of the archaeal domain.

Materials and Methods:

Sample Collection and Cultivation. Sediment and water samples were collected by B.P.H. from Obsidian Pool, Yellowstone National Park, USA, and ranged from 78^{o}-92^{o} C with a pH of ca. 6.5. The cultivation conditions for the Obsidian Pool enrichment culture were similar to those described previously (38). For details see SI *Text*.

Fluorescence *In-situ* Hybridization Analysis. FISH analysis was performed similar to described previously (39). Cy3-labeled, oligonucleotide probes KR515R (CCAGCCTTACCCTCCCCT) and KR565R (AGTATGCGTGGGAACCCCTC) provided optimal results. The hybridization solution containing 0.9 M NaCl, 0.5% SDS, 100 μg/ml

sheared herring sperm DNA, 0.02 M Tris-HCl (pH 7.2), and 20% formamide (v/v). The wash buffer containing 0.23 M NaCl, 0.1% (w/v) SDS, and 0.02 M Tris-Hcl (pH 7.2). For details see SI *Text*.

Electron Microscopy. Cell pellets were immediately fixed in a solution containing 2.5% glutaraldehyde (EM grade) in 20 mM sodium cacodylate buffer (pH 6.5). EM method details are provided in SI *Text*.

Cell Purification. Fermentor effluent was collected in sterile 2 l glass bottles. Washed cells were briefly exposed to 0.2% SDS (w/v) and then washed 3 times with PBS (pH 7.2). Cell suspensions were then filtered through 0.45 μm syringe filters (MILLEX HV, Millipore, Ireland) in 25 ml aliquots. The filtrate was centrifuged at 6000 rpm for 30 min. to collect the cells. See SI *Text* for detailed protocol.

Genome Sequencing and Assembly. Library construction, sequencing, and assembly were performed at the Joint Genome Institute, Walnut Creek, CA, USA. See SI *Text*.

Comparative Genomics and Phylogenetic Analyses. The predicted protein-coding genes were compared against those from other genomes available in the Integrated Microbial Genomes (IMG) analysis tool (52) and the National Center for Biotechnology Information database. Archaeal COGs were analyzed using the COGNITOR methods (19, 20, 53). An alignment of concatenated small and large subunit rRNA sequences (SSU+LSU rRNA) was constructed based on their conserved secondary structures and refined by hand. See SI *Text* for detailed information regarding phylogenetic analysis and tree construction.

This paper is dedicated to Carl Woese on the occasion of his 80th birthday. We thank Norman Pace for insightful comments regarding preparation of the final manuscript. We also thank the JGI production sequencing team; plus, Miriam Land and the Computational Biology Group at ORNL. Funding was provided to J.G.E and K.O.S. by Verenium (formerly Diversa) Corporation, San Diego, California, and The Deutsche Forschungsgemeinschaft. Support for genome sequencing and assembly was provided by the U.S. Dept. of Energy and the JGI Community Sequencing Program. Support to B.P.H. was provided by the Alexander von Humboldt Foundation and NSF no. MCB-0546865.

References:

1. Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 87:4576-4579.
2. Barns SM, Fundyga RE, Jeffries MW, Pace NR (1994) Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc Natl Acad Sci USA* 91:1609-1613.
3. Barns SM, Delwiche CF, Palmer JD, Pace NR (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc Natl Acad Sci USA* 93:9188-9193.

4. Auchtung TA, Takacs-Vesbach CD, Cavanaugh CM (2006) 16S rRNA phylogenetic investigation of the candidate division "Korarchaeota". *Appl Environ Microbiol* 72:5077-5082.
5. Takai K, Horikoshi K (1999) Genetic diversity of archaea in deep-sea hydrothermal vent environments. *Genetics* 152:1285-1297.
6. Inagaki F, *et al.* (2003) Microbial communities associated with geological horizons in coastal subseafloor sediments from the sea of okhotsk. *Appl Environ Microbiol* 69:7224-7235.
7. Waters E, *et al.* (2003) The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc Natl Acad Sci USA* 100:12984-12988.
8. Brochier C, Gribaldo S, Zivanovic Y, Confalonieri F, Forterre P (2005) Nanoarchaea: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to Thermococcales? *Genome Biol* 6:R42.
9. Reysenbach AL, Ehringer M, Hershberger K (2000) Microbial diversity at 83 degrees C in Calcite Springs, Yellowstone National Park: another environment where the Aquificales and "Korarchaeota" coexist. *Extremophiles* 4:61-67.
10. Hjorleifsdottir S, Skirnisdottir S, Hreggvidsson GO, Holst O, Kristjansson JK (2001) Species composition of cultivated and noncultivated bacteria from short filaments in an Icelandic hot spring at 88 degrees C. *Microb Ecol* 42:117-125.
11. Skirnisdottir S, *et al.*(2000) Influence of sulfide and temperature on species composition and community structure of hot spring microbial mats. *Appl Environ Microbiol* 66:2835-2841.
12. Takai K, Yoshihiko S (1999) A molecular view of archaeal diversity in marine and terrestrial hot water environments. *FEMS Microbiol Ecol* 28:177-188.
13. Marteinsson VT, *et al.* (2001) Discovery and description of giant submarine smectite cones on the seafloor in Eyjafjordur, northern Iceland, and a novel thermal microbial habitat. *Appl Environ Microbiol* 67:827-833.
14. Schrenk MO, Kelley DS, Delaney JR, Baross JA (2003) Incidence and diversity of microorganisms within the walls of an active deep-sea sulfide chimney. *Appl Environ Microbiol* 69:3580-3592.
15. Nercessian O, Reysenbach AL, Prieur D, Jeanthon C (2003) Archaeal diversity associated with in situ samplers deployed on hydrothermal vents on the East Pacific Rise (13 degrees N). *Environ Microbiol* 5:492-502.
16. Spear JR, Walker JJ, McCollom TM, Pace NR (2005) Hydrogen and bioenergetics in the Yellowstone geothermal ecosystem. *Proc Natl Acad Sci USA* 102:2555-2560.
17. Teske A, *et al.* (2002) Microbial diversity of hydrothermal sediments in the Guaymas Basin: evidence for anaerobic methanotrophic communities. *Appl Environ Microbiol* 68:1994-2007.
18. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955-964.
19. Makarova KS, Sorokin AV, Novichkov PS, Wolf YI, Koonin EV (2007) Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea. *Biol Direct* 2:33.
20. Tatusov RL, *et al.* (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41-54.
21. Sapra R, Bagramyan K, Adams MWW (2003) A simple energy-conserving system: Proton reduction coupled to proton translocation. *Proc Natl Acad Sci USA* 100:7545-7550.
22. Lindahl PA, Graham DE (2007) in *Nickel and Its Surprising Impact in Nature*, eds Sigel A, Sigel H, Sigel RKO (John Wiley & Sons, Chichester, UK), pp. 357-416.

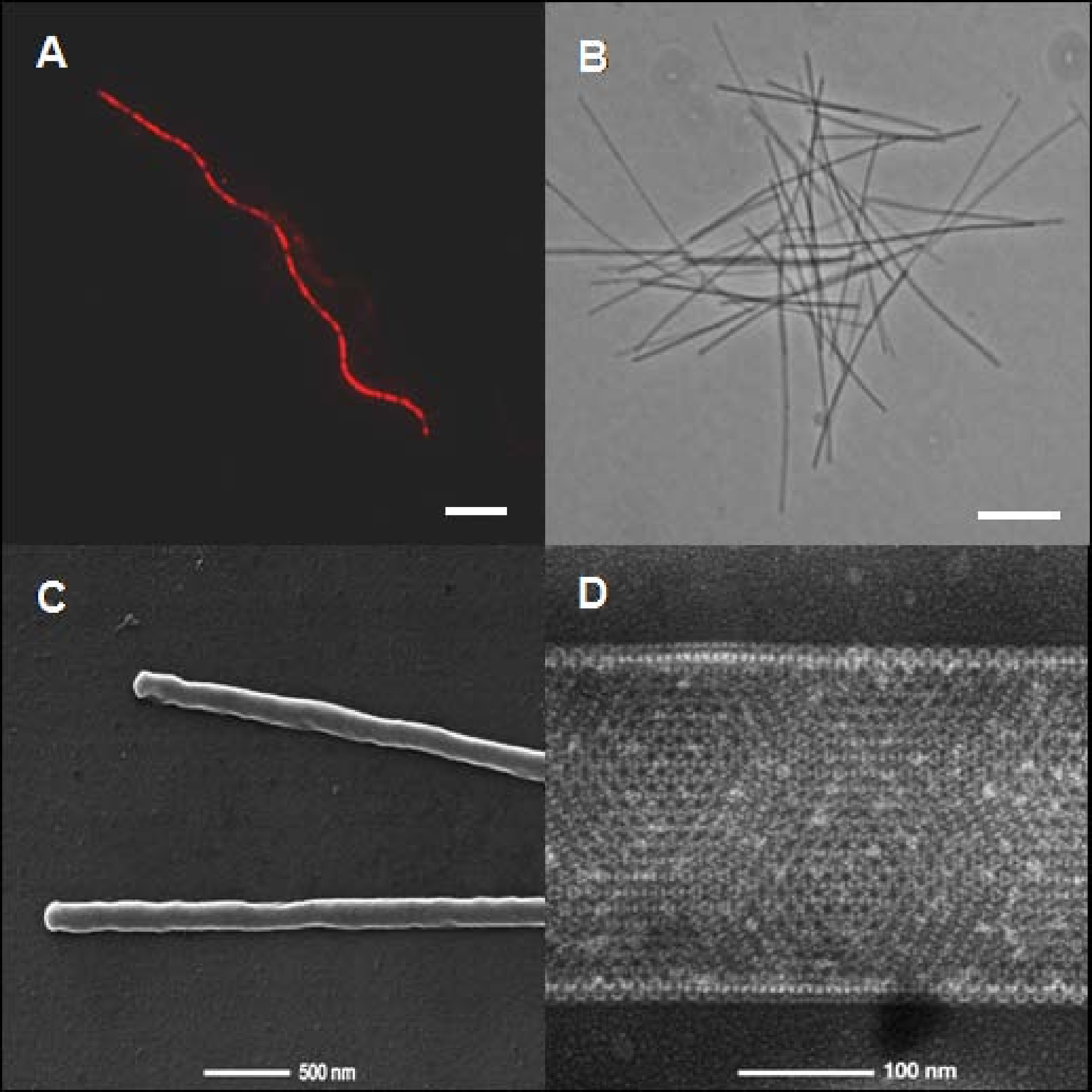
23. Orita I, *et al.*, (2006) The Ribulose Monophosphate Pathway Substitutes for the Missing Pentose Phosphate Pathway in the Archaeon *Thermococcus kodakaraensis*. *J. Bacteriol.* 188:4698-4704.
24. Anderson I, *et al.* (2008) Genome sequence of *Thermofilum pendens* reveals an exceptional loss of biosynthetic pathways without genome reduction. *J Bacteriol*. E-pub
25. Marinsek N, *et al.* (2006) GINS, a central nexus in the archaeal DNA replication fork. *EMBO Rep* 7:539-545.
26. Iyer LM, Makarova KS, Koonin EV, Aravind L (2004) Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. *Nucleic Acids Res* 32:5260-5279.
27. Vaughan S, Wickstead B, Gull K, Addinall SG (2004) Molecular evolution of FtsZ protein sequences encoded within the genomes of archaea, bacteria, and eukaryota. *J Mol Evol* 58:19-29.
28. Faguy DM, Doolittle WF (1998) Cytoskeletal proteins: the evolution of cell division. *Curr Biol* 8:R338-341.
29. Poplawski A, Gullbrand B, Bernander R (2000) The ftsZ gene of *Haloferax mediterranei*: sequence, conserved gene order, and visualization of the FtsZ ring. *Gene* 242:357-367.
30. Ye H, *et al.* (2004) Crystal structure of the putative adapter protein MTH1859. *J Struct Biol* 148:251-256.
31. Koonin EV, Makarova KS, Elkins JG (2007) Orthologs of the small RPB8 subunit of the eukaryotic RNA polymerases are conserved in hyperthermophilic Crenarchaeota and "Korarchaeota". *Biology Direct* 2.
32. Cann IK, Ishino S, Nomura N, Sako Y, Ishino Y (1999) Two family B DNA polymerases from *Aeropyrum pernix*, an aerobic hyperthermophilic crenarchaeote. *J Bacteriol* 181:5984-5992.
33. Lecompte O, Ripp R, Thierry JC, Moras D, Poch O (2002) Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res* 30:5382-5390.
34. Marck C, Grosjean H (2002) tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA* 8:1189-1232.
35. Parrish J, *et al.* (2001) Mitochondrial endonuclease G is important for apoptosis in *C. elegans*. *Nature* 412:90-94.
36. Feng L, Sheppard K, Tumbula-Hansen D, Söll D (2005) Gln-tRNAGln Formation from Glu-tRNAGln Requires Cooperation of an Asparaginase and a Glu-tRNAGln Kinase. *J Biol Chem* 280:8150-8155.
37. Kim HS, Vothknecht UC, Hedderich R, Celic I, Söll D (1998) Sequence divergence of seryl-tRNA synthetases in archaea. *J Bacteriol* 180:6446-6449.
38. Burggraf S, Heyder P, Eis N (1997) A pivotal Archaea group. *Nature* 385:780.
39. Burggraf S, *et al.* (1994) Identifying members of the domain Archaea with rRNA-targeted oligonucleotide probes. *Appl Environ Microbiol* 60:3112-3119.
40. Kelly RM, Adams MW (1994) Metabolism in hyperthermophilic microorganisms. *Antonie Van Leeuwenhoek* 66:247-270.
41. Ma K, Schicho RN, Kelly RM, Adams MW (1993) Hydrogenase of the hyperthermophile *Pyrococcus furiosus* is an elemental sulfur reductase or sulfhydrogenase: evidence for a sulfur-reducing hydrogenase ancestor. *Proc Natl Acad Sci USA* 90:5341-5344.

42. Mai X, Adams MW (1994) Indolepyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*. A new enzyme involved in peptide fermentation. *J Biol Chem* 269:16726-16732.
43. Blamey JM, Adams MW (1993) Purification and characterization of pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Biochim Biophys Acta* 1161:19-27.
44. Schut GJ, Brehm SD, Datta S, Adams MW (2003) Whole-genome DNA microarray analysis of a hyperthermophile and an archaeon: *Pyrococcus furiosus* grown on carbohydrates or peptides. *J Bacteriol* 185:3935-3947.
45. Huber R, Huber H, Stetter KO (2000) Towards the ecology of hyperthermophiles: biotopes, new isolation strategies and novel metabolic properties. *FEMS Microbiol Rev* 24:615-623.
46. Robertson CE, Harris JK, Spear JR, Pace NR (2005) Phylogenetic diversity and ecology of environmental Archaea. *Curr Opin Microbiol* 8:638-642.
47. Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P (2008) Mesophilic crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol* 6:245-252.
48. Hallam SJ, *et al.* (2006) Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc Natl Acad Sci USA* 103:18296-18301.
49. Cubonova L, Sandman K, Hallam SJ, Delong EF, Reeve JN (2005) Histones in crenarchaea. *J Bacteriol* 187:5482-5485.
50. Sandman K, Reeve JN (2006) Archaeal histones and the origin of the histone fold. *Curr Opin Microbiol* 9:520-525.
51. Lopez-Garcia P, Brochier C, Moreira D, Rodriguez-Valera F (2004) Comparative analysis of a genome fragment of an uncultivated mesopelagic crenarchaeote reveals multiple horizontal gene transfers. *Environ Microbiol* 6:19-34.
52. Markowitz VM, *et al.* (2006) The integrated microbial genomes (IMG) system. *Nucleic Acids Res* 34:D344-348.
53. Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. *Science* 278:631-637.

Figure legends:

Fig. 1. Microscopy of "*Ca*. K. cryptofilum". (*A*) FISH analysis with *Korarchaeota*-specific Cy3-labeled oligonucleotide probes KR515R/KR565R. The undulated cell shape results from drying of the specimen on gelatin coated slides prior to hybridization. Scale bar represents 5 μm. (*B*) Phase contrast image of korarchaeal filaments following physical enrichment. Scale bar represents 5 μm. (*C*), Scanning electron micrograph of purified cells. (*D*) Transmission electron micrograph after negative staining with uranyl acetate displaying the paracrystalline S-layer. Cells are flattened which increases their apparent thickness.

Fig. 2. Phylogenetic analysis of "*Ca*. K. cryptofilum". (*A*) Maximum likelihood phylogenetic tree of combined (SSU+LSU) rRNAs rooted with corresponding bacterial sequences. Numbers at the nodes indicate bootstrap support. (*B*) Archaeal phylogeny based on translation elongation factor 2 (EF2) proteins rooted with bacterial homologs. The numbers indicate bootstrap support for PhyML/consensus posterior probability (Phyloblast), an asterisk indicates <50 bootstrap support. Where both values were <50, the branch was collapsed. (*C*) Maximum likelihood tree made from aligned sequences of 33 universally conserved ribosomal proteins and the 3 largest RNA polymerase subunits, RpoA, RpoB, and RpoD. Bootstrap support numbers are given at the nodes as a percentage (n=10000). Scale bars represent the average number of substitutions per residue.


A
B
C
D
500 nm
100 nm

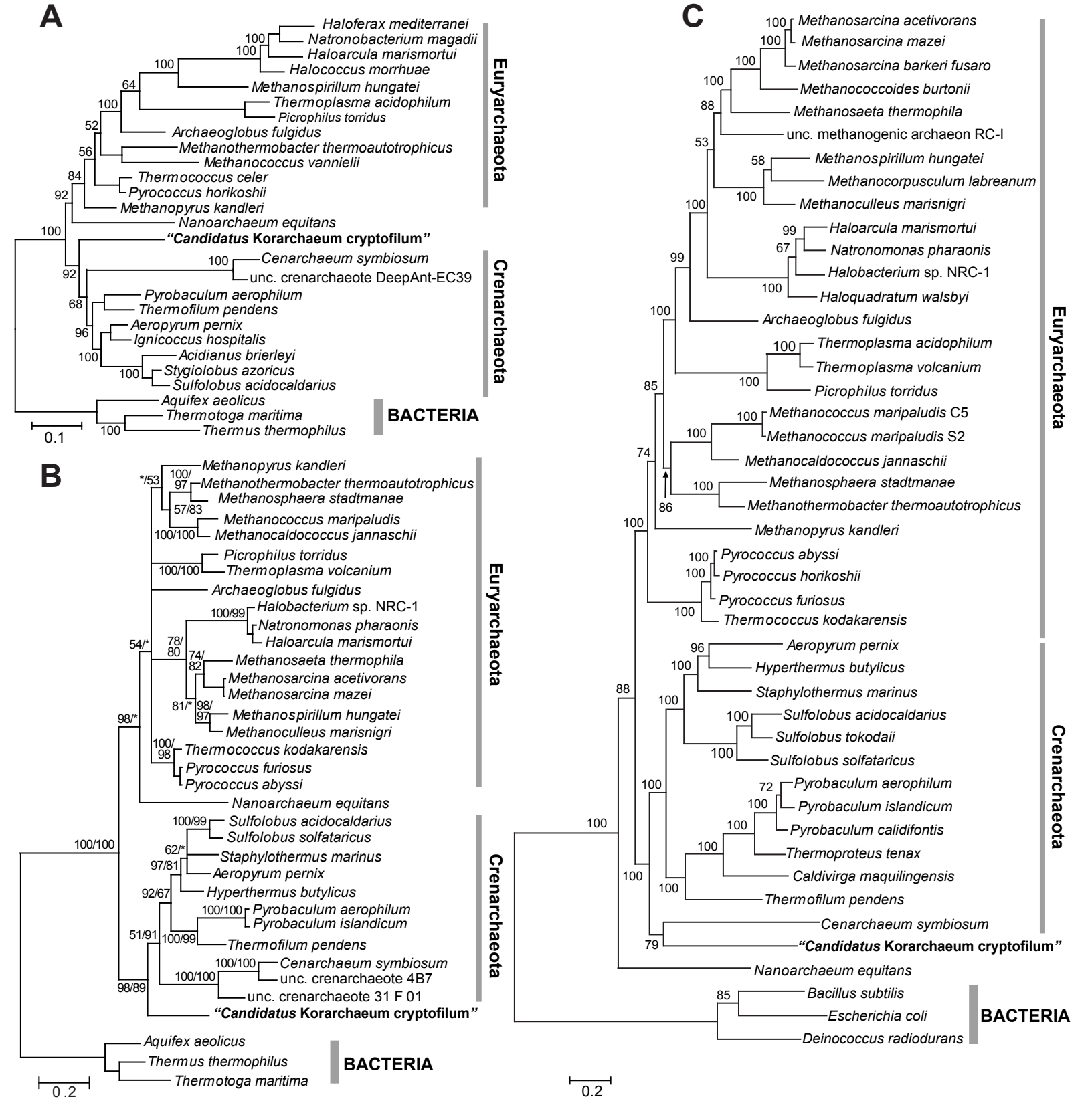

A
Haloferax mediterranei
Natronobacterium magadii
Haloarcula marismortui
Halococcus morrhuae
Methanospirillum hungatei
Thermoplasma acidophilum
Picrophilus torridus
Archaeoglobus fulgidus
Methanothermobacter thermoautotrophicus
Methanococcus vannielii
Thermococcus celer
Pyrococcus horikoshii
Methanopyrus kandleri
Nanoarchaeum equitans
"Candidatus Korarchaeum cryptofilum"
Cenarchaeum symbiosum
unc. crenarchaeote DeepAnt-EC39
Pyrobaculum aerophilum
Thermofilum pendens
Aeropyrum pernix
Ignicoccus hospitalis
Acidianus brierleyi
Stygiolobus azoricus
Sulfolobus acidocaldarius
Aquifex aeolicus
Thermotoga maritima
Thermus thermophilus
0.1
Euryarchaeota
Crenarchaeota
BACTERIA
B
Methanopyrus kandleri
Methanothermobacter thermoautotrophicus
Methanosphaera stadtmanae
Methanococcus maripaludis
Methanocaldococcus jannaschii
Picrophilus torridus
Thermoplasma volcanium
Archaeoglobus fulgidus
Halobacterium sp. NRC-1
Natronomonas pharaonis
Haloarcula marismortui
Methanosaeta thermophila
Methanosarcina acetivorans
Methanosarcina mazei
Methanospirillum hungatei
Methanoculleus marisnigri
Thermococcus kodakarensis
Pyrococcus furiosus
Pyrococcus abyssi
Nanoarchaeum equitans
Sulfolobus acidocaldarius
Sulfolobus solfataricus
Staphylothermus marinus
Aeropyrum pernix
Hyperthermus butylicus
Pyrobaculum aerophilum
Pyrobaculum islandicum
Thermofilum pendens
Cenarchaeum symbiosum
unc. crenarchaeote 4B7
unc. crenarchaeote 31 F 01
"Candidatus Korarchaeum cryptofilum"
Aquifex aeolicus
Thermus thermophilus
Thermotoga maritima
0.2
Euryarchaeota
Crenarchaeota
BACTERIA
C
Methanosarcina acetivorans
Methanosarcina mazei
Methanosarcina barkeri fusaro
Methanococcoides burtonii
Methanosaeta thermophila
unc. methanogenic archaeon RC-I
Methanospirillum hungatei
Methanocorpusculum labreanum
Methanoculleus marisnigri
Haloarcula marismortui
Natronomonas pharaonis
Halobacterium sp. NRC-1
Haloquadratum walsbyi
Archaeoglobus fulgidus
Thermoplasma acidophilum
Thermoplasma volcanium
Picrophilus torridus
Methanococcus maripaludis C5
Methanococcus maripaludis S2
Methanocaldococcus jannaschii
Methanosphaera stadtmanae
Methanothermobacter thermoautotrophicus
Methanopyrus kandleri
Pyrococcus abyssi
Pyrococcus horikoshii
Pyrococcus furiosus
Thermococcus kodakarensis
Aeropyrum pernix
Hyperthermus butylicus
Staphylothermus marinus
Sulfolobus acidocaldarius
Sulfolobus tokodaii
Sulfolobus solfataricus
Pyrobaculum aerophilum
Pyrobaculum islandicum
Pyrobaculum calidifontis
Thermoproteus tenax
Caldivirga maquilingensis
Thermofilum pendens
Cenarchaeum symbiosum
"Candidatus Korarchaeum cryptofilum"
Nanoarchaeum equitans
Bacillus subtilis
Escherichia coli
Deinococcus radiodurans
0.2
Euryarchaeota
Crenarchaeota
BACTERIA

Table 1. General features of the "*Ca*. K. cryptofilum" genome

Total number of bases	1590757
Coding density (%)	89
G + C content (%)	49
Total number of predicted genes	1661
Protein coding genes	1602
Average ORF length (bp)	870
rRNA genes*	3
tRNA genes	45
Genes assigned to COGs	1401
Genes assigned to arCOGs	1382
Genes with function prediction	998
Genes with biochemical prediction only	246
Genes with unknown function or activity	358

*16S, 23S, and 5S rRNA

Table 2. Crenarchaeal and euryarchaeal arCOGs in *"Ca.* K. cryptofilum"

arCOG	Cat.*	Function	Eu‡	Cr§
04447	L	DNA polymerase II, large subunit	27	0
04455	L	DNA polymerase II, small subunit	26	0
00872	L	ERCC4-like helicase	26	0
02610	L	Rec8/ScpA/Scc1-like protein	24	0
02258	L	subunit of RPA complex	20	0
00371	D	Chromosome segregation ATPase, SMC	24	0
02201	D	Cell division GTPase FtsZ	26	0
01013	J	Protein with L13E-like domain	0	11
04327	J	Ribosomal protein S25	0	13
04293	J	Ribosomal protein S30	0	13
04305	J	Ribosomal protein S26	0	13
04271	K	RNA polymerase, subunit RPB8	0	12
00393	K	Membrane-associated transcriptional regulator	0	9

*COG functional categories: L, Replication, recombination and repair; D, Cell cycle control, cell division, chromosome partitioning: J, Translation, ribosomal structure and biogenesis; K, Transcription. ‡Number of euryarchaeal (Eu) genomes containing that arCOG (out of 27 total). §Number of crenarchaeal (Cr) genomes containing that arCOG (13 total).

Supporting Text

Results:

Cultivation and Cell Identification. Members of the Crenarchaeota were represented by clone types pOPF_10, pOPF_07, and pOPF_12 (OPF = Obsidian Pool Fermentor) which shared 98% similarity to *Thermosphaera aggregans* strain M11TL^T (1); 98% similarity to *Thermofilum pendens* S10TFL (1); and 98% similarity to an uncultured *Pyrobaculum* sp. HVerd019N respectively (Kvist, 2006; unpublished). Two closely related organisms within the Euryarchaeota were designated pOPF_1 and pOPF_3 and were 98% and 99% similar, respectively, to an uncultivated member of the Archaeoglobales from Obsidian Pool designated OPPDO15 (2). A single member of the *Korarchaeota*, represented by the 16S rDNA clone pOPF_08, was supported by the enrichment culture. The 16S sequence of this "korarchaeote" is 99% similar to both the pJP27 sequence type from Obsidian Pool, YNP (3), and the pAB5 sequence from Calcite Springs, YNP (4) (see SI Fig. 3).

FISH Analysis. Cy3-labeled, oligonucleotide probes (*5'-3'*) KR515R (CCAGCCTTACCCTCCCCT) and KR565R (AGTATGCGTGGGAACCCCTC) allowed visualization of the korarchaeal filamentous cells in single probe hybridizations but provided the best results when applied in combination. A crenarchaeota-specific probe, CREN499R (CCARNCTTGCCCCCCGCT) labelled with Alexa488, failed to hybridize to KR515R/KR565R positive filaments however showed bright fluorescence when pure cultures of *Thermofilum sp.* S10TFL were used as a positive control (SI Figs. 4*A* and 4*B*). Probes KR515R and KR565R did not hybridize to pure cultures of *Thermofilum sp.* S10TFL. Domain specific probes ARC915 (5'-GTGCTCCCCCGCCAATTCCT-3') (5) and EUB338 (5'-GCTGCCTCCCGTAGGAGT-3') (6) also did not hybridize to the *Korarchaeota* positive filaments (data not shown). Addition of sheared fish sperm DNA to the hybridization buffer reduced background fluorescence. The SDS concentration in the hybridization buffer had to be increased to 0.5% (w/v) to allow sufficient probe penetration.

Sequencing and Assembly. Sample sequencing was used to determine which libraries represented the highest number of korarchaeal genomic sequences. The fraction of end-reads that aligned with the 84.7 kb korarchaeal genomic contig (SI Fig. 6) was determined. Libraries BFPP and especially BHXI yielded the highest proportion of sequences that overlapped with the korarchaeal 84.7 kb contig and were selected for further sequencing. A total of 23,000 and 11,520 sequencing reads were produced from the BHXI and BFPP libraries respectively. The average read length was 800 +/- 113 bp for the small-insert BHXI library and 625 +/- 170 bp for the fosmid based BFPP library. A high number of reads (2292) from the BFPP library were of insufficient length and were not used for further analysis. An additional 6184 sequencing reads from the BFPP library were also determined to be from other Obsidian Pool enrichment culture organisms (based on %GC analysis) and were therefore rejected. Attempts at assembling the "*Candidatus* Korarchaeum cryptofilum" genome were made using the 23000 quality sequencing reads from the BHXI library combined with 3044 reads from the large-insert library. From the assembly data, it was evident that the original DNA sample contained a dominant genome represented by the pOPF_08 clone type and at least two other genomic variants in low concentration. Since the dominate genome was sufficiently

enriched relative to other variants (approximately 10-fold), genomic heterogeneity did not prevent the assembly of the complete genome of the pOPF_08 organism.

Amino Acid Biosynthesis. Although "*Ca.* K. cryptofilum" appears to be a proficient peptide degrader, it has an extensive set of amino acid biosynthesis enzymes. The predicted repertoire includes a canonical branched chain amino acid pathway and redundant threonine dehydratase and citramalate synthases to produce 2-oxobutyrate, the precursor for isoleucine biosynthesis. Similar to crenarchaeotes and the *Thermococcales*, "*Ca.* K. cryptofilum" appears to use the homocitrate pathway and genes of the *lysYZWXJK* operon to produce lysine (7). However, it also encodes *dapA*, *dapE*, *dapC,* and *lysA* homologs from the diaminopimelate pathway of lysine synthesis although *dapB*, *dapD,* and *dapF* genes are missing. Also missing are histidine and tryptophan biosynthesis genes, although the prephenate pathway to phenylalanine and tyrosine is complete. Unlike the known Crenarchaeota, the korarchaeote is predicted to use the methanogenic version of chorismate synthesis (8). Common pathways are available to produce arginine, proline, threonine, and serine can be made through the phosphorylating pathway.

Mobile Elements. The "*Ca.* K. cryptofilum" genome contains 6 genes (Kcr_0344, Kcr_0272, Kcr_1558, Kcr_0611, Kcr_1387, Kcr_0856) that encode OrfB of the IS605 family of transposases (9) which are widespread throughout the Archaea (10). Kcr_1558 and Kcr_1387 share 87% identity and both alleles are clustered with predicted ISC1913-like elements similar to those found in *Sulfolobus solfataricus* P2 (11). Kcr_0344 and Kcr_0272 are also nearly identical to one another and are closely related to IS605 sequences found in hyperthermophilic crenarchaeotes as well as transposable elements carried by conjugal plasmids (12).

Tree Compatibility. Analysis of concatenated protein alignments hinges on the assumption that the true phylogenies of all genes involved are the same and the observed differences between trees reconstructed from individual genes represent noise. Thus, ideally, genes must be tested for compatibility of their individual phylogenies before concatenation. Practically, when dealing with large set of genes, such as >30 r-proteins and 6 RNAP subunits, complete rigorous testing becomes unfeasible. Thus we employed the following scheme:

The 6 RNAP subunit alignments were concatenated, and an ML phylogenetic tree was reconstructed from the concatenated alignment. Each original alignment was tested for compatibility with the topology of this tree using the Approximately Unbiased (AU) test (13) as implemented in the TreeFinder program (14). Three subunits (E', F and N) were found to be incompatible with the concatenated tree and, accordingly, were rejected; three (β and two chains of β') passed the test (data not shown). The concatenated alignment of these three subunits was used in all further analyses.

Individual ribosomal proteins are generally too short to be used for a meaningful phylogenetic reconstruction individually, so their alignments were concatenated without detailed testing. We further assessed the compatibility of phylogenetic tree topologies between the concatenated r-proteins and concatenated RNAP subunits. First, *M. kandleri* sequences were removed from both datasets as it has been shown earlier that translational and transcriptional machineries of this archaeon appear to be of different origins (15). Maximum likelihood phylogenetic trees were

reconstructed for the concatenated r-proteins, concatenated RNAP subunits, and the two protein sets concatenated together. All three reconstructions placed "*Ca.* K. cryptofilum", *C. symbiosum* and *N. equitans* into slightly different deep positions in the archaeal tree. Each of the three reconstructed phylogenies was reduced to an "essential constraint" tree which preserves the deep branching order of "*Ca.* K. cryptofilum", *C. symbiosum*, *N. equitans*, Crenarchaeota clade, the Euryarchaeota clade, and the Bacteria clade but reduces the rest of the tree to a polytomies. Multifurcations in each of the three constraint trees were resolved separately for the concatenated r-proteins and concatenated RNAP subunits. Each of the two protein sets was tested for compatibility with two alternative topologies against its native topology using the AU test.

The tree topologies of concatenated r-proteins and concatenated RNAP subunits were mutually incompatible (AU p-value is <0.0001 for both cross comparisons). However, each topology was compatible with the topology of the joint tree that was obtained from the concatenation of both alignments (AU p-value of 0.44 for the r-proteins and 0.09 for the RNAP subunits). This concatenation provides the "middle ground" topology which is compatible with both original components.

For compatibility testing, the trees are given below in the Newick format or can be represented as rooted or unrooted trees using the TreeView program (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html)

Rib+Rpo:

(((((((Aerpe:0.36773495,Hypbu:0.21297609):0.058441252,Stama:0.27448517):0.066622257,((Sulac:0.15015128,Sulto:0.10456212):0.069624523,Sulso:0.1509329):0.2630078):0.08883196,((Calma:0.3253224,(((Pyrae:0.054856231,Pyris:0.067626872):0.025078936,Pyrca:0.059170292):0.10683526,Thete:0.15772165):0.15116257):0.19498017,Thepe:0.3837574):0.091744444):0.084763622,(Censy:0.76822737,Korar:0.6614682):0.069591385):0.072259578,(((Bacsu:0.33698426,Escco:0.44608078):0.10326424,Deira:0.41947042):1.5273464,Naneq:0.64862503):0.086190568):0.066019763,(((Pyrab:0.012910551,Pyrho:0.022003269):0.01624607,Pyrfu:0.023925165):0.052710879,Theko:0.081936192):0.25734593):0.055934894,((Arcfu:0.34082045,((((((Halma:0.11561483,Natph:0.1182512):0.038811422,Halsp:0.13959571):0.038397559,Halwa:0.14293013):0.40280754,Uncme:0.28868764):0.040151522,((((Metac:0.019253391,Metma:0.02656954):0.019116731,Metba:0.040855086):0.1295261,Metbu:0.18549369):0.1495346,Metsa:0.2967868):0.049075929):0.039992591,(Metcu:0.16004155,
(Methu:0.19810155,Metla:0.26673314):0.036847236):0.24328027):0.097104881):0.070349984,(Picto:0.21986978,(Theac:0.070725281,Thevo:0.069452574):0.15861092):0.44427765):0.060677377,((Metja:0.14230929,(MetmC:0.015643543,Metmp:0.014565799):0.25379213):0.19458358,(Metst:0.23903843,Metth:0.1303759):0.23852987):0.041906053);

Rib:

(((((((Aerpe:0.31980804,Hypbu:0.20615681):0.055498529,Stama:0.25718752):0.066253714,((Sulac:0.16576925,Sulto:0.11668129):0.077594061,Sulso:0.16282142):0.28501161):0.093247323,((C

alma:0.33193714,(((Pyrae:0.057120572,Pyris:0.07077073):0.026296476,Pyrca:0.066530324):0.10905667,Thete:0.17545736):0.1438332):0.18642503,Thepe:0.38173286):0.099866176):0.085274407,(Censy:0.82794032,Korar:0.56696967):0.068304266):0.07088327,Naneq:0.68565168):0.044133173,((Bacsu:0.33527378,Escco:0.45321524):0.098950092,Deira:0.41249964):1.3966304):0.033603488,(((Arcfu:0.37551694,(((((Halma:0.13342765,Natph:0.1326585):0.043048517,Halsp:0.15745755):0.043228326,Halwa:0.15991492):0.42703979,(Metcu:0.18289195,(Methu:0.22837977,Metla:0.29610791):0.043372113):0.21833222):0.035515818,(((((Metac:0.021885312,Metma:0.027189481):0.020552123,Metba:0.047812763):0.13917693,Metbu:0.19809341):0.16309816,Metsa:0.34037388):0.049836532,Uncme:0.34735559):0.03902647):0.10982473):0.060987418,(Picto:0.23763709,(Theac:0.084552
977,Thevo:0.085595389):0.18410383):0.45236938):0.058668401,((Metja:0.14052579,(MetmC:0.015882315,Metmp:0.015780284):0.26479904):0.18936641,(Metst:0.25169702,Metth:0.13607037):0.22555101):0.039398826):0.057778373,(((Pyrab:0.014396699,Pyrho:0.024345608):0.014770393,Pyrfu:0.020655354):0.056742521,Theko:0.0853559):0.24336287);

Rpo:

(((((((Aerpe:0.52456895,Hypbu:0.24236839):0.069374313,Stama:0.33298331):0.068457671,((Sulac:0.11654009,Sulto:0.077091543):0.052514386,Sulso:0.12574242):0.21556208):0.07859586,((Calma:0.31876506,(((Pyrae:0.050389173,Pyris:0.062111294):0.022680793,Pyrca:0.042864628):0.10570598,Thete:0.11873635):0.17435016):0.2250776,Thepe:0.40257348):0.072575108):0.077953824,Censy:0.71857094):0.064825184,(((Bacsu:0.34625103,Escco:0.4453692):0.10840656,Deira:0.46139837):2.2483465,(Korar:0.90258868,Naneq:0.75175588):0.10023721):0.060014185):0.072094977,(((Pyrab:0.0096322551,Pyrho:0.016736461):0.019982278,Pyrfu:0.033581242):0.043467586,Theko:0.076615985):0.3032898):0.054780751,((Arcfu:0.27501647,(((Halma:0.076647508,(Halsp:0.099864933,Halwa:0.12959176):0.030925598):0.02240101,Natph:0.067741209):0.36352804,((((Metac:0.013346087,Metma:0.025591969):0.015592053,Metba:0.024615296):0.10901293,Metbu:0.16095282):0.12200913,Metsa:0.20025675):0.038264588,(((Metcu:0.12050679,Methu:0.13685102):0.032523153,M
etla:0.18647727):0.26028804,Uncme:0.21766027):0.038322895):0.047884765):0.073732157):0.094247225,(Picto:0.18714592,(Theac:0.039550499,Thevo:0.033687756):0.10319597):0.43834352):0.063219904,((Metja:0.15386209,(MetmC:0.016300163,Metmp:0.010996696):0.23172211):0.21688718,(Metst:0.21362919,Metth:0.12149914):0.28171906):0.050458485);

Supporting Materials and Methods:

Sample Collection and Cultivation. A variety of sediment and water samples were collected from Obsidian Pool, Yellowstone National Park, Wyoming, USA, and ranged from 78^{o}-92^{o} C and had a pH of ca. 6.5. Similar samples collected near the same time our samples were taken (within days) were subjected to a detailed geochemical analysis by Schock et al. (16). The enrichment culture was established by adding several reduced and nonreduced samples (approximately 400 ml) of sediment and spring water to a double-walled glass vessel with a total volume of ca. 800 ml. The Inoculation was performed inside a Coy anaerobic tent (Coy Laboratories, Grass Lake, MI, USA) to avoid exposure to air. Heated glycerol was pumped through the outer chamber of the

vessel to maintain a constant temperature of 85° C. A gas mixture consisting of N_2/CO_2 (80:20) was constantly bubbled through the culture at a rate of 20 ml/min. The anaerobic growth medium consisted of a modified Allen's formulation (see below) (17) and was continuously added to the vessel at a rate of 8.0 ml/h using a Gilson Minipuls2 peristaltic pump (Gilson Inc., Middleton, WI, USA). Culture effluent was allowed to overflow through a side port in the cultivation vessel and into a sterile, anaerobic 2 l glass media bottle.

Obsidian Pool fermentor (OPF) medium (modified from Allen, 1959)

$(NH_4)_2SO_4$	1.30 g
KH_2PO_4	0.28 g
$MgSO_4$ x 7 H_2O	0.25 g
$CaSO_4$ x 2 H_2O	0.17 g
$CaCl_2$ x 2 H_2O	0.07 g
$FeCl_3$ x 6 H_2O	0.02 g
Na_2SO_4	0.07 g
KNO_3	0.10 g
Peptone	0.50 g
Yeast Extract	0.10 g
$Na_2S_2O_3$	0.78 g
Na_2S x 6 H_2O	0.13 g
$MnCl_2$ x 4 H_2O (10 mg/ml)	180 µl
$Na_2B_4O_7$ x 10 H_2O (25 mg/ml)	180 µl
$ZnSO_4$ x 7 H_2O (10 mg/ml)	22 µl
$CuCl_2$ x 2 H_2O (10 mg/ml)	5 µl
Na_2MoO_4 x 2 H_2O (1 mg/ml)	30 µl
$VOSO_4$ x 5 H_2O (1 mg/ml)	30 µl
$CoSO_4$ x 7 H_2O (1 mg/ml)	10 µl
$LiCl/Na_2WO_4/NaSeO_3/Ni(NH_4)_2(SO_4)$ (1 mg/ml each)	10 µl
Wolfe's Vitamin Solution (1000X)	1.0 ml
Distilled H2O	1000 ml

Fluorescent *In-situ* Hybridization Analysis. Fresh samples were harvested directly from the Obsidian Pool enrichment culture and washed 3 times with 1 ml of phosphate buffered saline (PBS; pH 7.2). The cells were then spotted onto precleaned, gelatin coated microscope slides in 8 µl aliquots and dried by placing the slides onto a 50°C heat block. Ten microliters of fixative solution [3% paraformaldehyde (w/v) in PBS (pH 7.2)] was placed over the dried cells and the samples were then incubated either at RT for 2 h or overnight at 4°C. The fixed cells were washed thoroughly by rinsing the slides with PBS (pH 7.2) followed by an ethanol dehydration series by submerging the slides into a 50%, 70%, and 100% ethanol (v/v) bath for 3 min each. The slides

were then dried by applying a gentle stream of compressed air. A hybridization chamber was prepared by saturating a folded paper towel with hybridization buffer and placing the towel in a 50 ml conical centrifuge tube. A volume of 8 µl of hybridization solution containing 0.9 M NaCl, 0.01%-0.5% (w/v) sodium dodecyl sulfate (SDS), 100 µg/ml sheared herring sperm DNA, 0.02 M Tris-HCl (pH 7.2), and 0-40% formamide (v/v) were applied to each cell spot and the slides were placed in the hybridization chamber and incubated at 46^{o} C for 15 min. After the prehybridization step, 100 ng of the appropriate fluorescently labeled, oligonucleotide probe was added to the hybridization buffer and the slides were incubated at 46^{o}C for 3 h. Excess probe was washed from the slide by gently pipetting 2 ml pre-warmed wash buffer containing 0.9M-0.06 M NaCl (depending of formamide concentration in hybridization buffer), 0.1% (w/v) SDS, and 0.02 M Tris-Hcl (pH 7.2) over the cells. The entire slide was then immersed in pre-warmed wash buffer and placed in a 48^{o} C water bath for 15 min. The slide was then rinsed with ca. 300 ml of deionized H_2O and dried by applying a gentle stream of compressed air. A drop of Prolong Gold mounting solution containing 4',6-Diamidino-2-phenylindole (DAPI) (Molecular Probes, San Diego, CA, USA) was applied to each cell spot followed by a 25 x 25 mm coverslip. All microscopic analysis was performed with a Zeiss Axioplan 2 microscope (Jena, Germany) with 100 W mercury arc and halogen illumination sources. Phase contrast images were collected using either 63X, or 100X Plan Neofluar oil immersion objectives with 1.4 and 1.3 numerical apertures respectively. Images were collected using a Zeiss Axiocam digital camera and processed with Zeiss Axiovision software.

Electron Microscopy. *Korarchaeota* cell pellets were fixed in a solution containing 2.5% glutaraldehyde (EM grade) in 20 mM sodium cacodylate buffer (pH 6.5). For scanning electron microscopy drops of the fixed sample were placed onto glass slides, covered with a cover slip and rapidly frozen with liquid nitrogen. The cover slip was removed with a razor blade and the slide was immediately transferred into fixative buffer, postfixed with osmium tetroxide, dehydrated in a graded series of acetone solutions and critical-point dried from liquid CO_2, mounted on stubs, and coated with 3 nm platinum with a magnetron sputter coater. The specimens were examined with a Hitachi S-4100 field emission scanning electron microscope. For negative staining a drop of the sample at appropriate dilution was placed on a 400 mesh carbon-coated copper grid, freshly hydrophilisized by glow discharge. After incubation for 2 min, the drop was quickly removed with a pasteur pipette and the grid was air dried. The grid was stained with 2% uranium acetate and 0.01% glucose. Micrographs were taken with an EM 912 electron microscope (Zeiss) equipped with an integrated OMEGA energy filter operated in the zero loss mode.

Cell Purification. Culture effluent from the Obsidian Pool fermentor was collected anaerobically in sterile 2 l glass bottles. The cells were pelleted by centrifugation at 6000 rpm in a Beckman JLA-10.500 rotor for 20 min at 25^{o}C in 500 ml centrifuge bottles (Nalgene, Rochester, NY, USA). The cells were resuspended in a total volume of 200 ml of PBS (pH 7.2) and then divided into 4 x 50 ml aliquots in 50 ml conical centrifuge tubes (BD Falcon, Bedford, MA, USA). The cells were centrifuged again at 6000 rpm for 20 min in a Beckman JA-12 rotor at 25^{o}C and the supernatant was discarded. Each pellet was resuspended again in 50 ml of PBS (pH 7.2) by vortexing. After the cell pellets were completely resuspended, 0.5 ml of a 20% (w/v) SDS solution [final conc. 0.2% SDS (w/v)] was added to each tube and immediately mixed by inverting the tube several times. The tubes were centrifuged again at 6000 rpm for 20 min in a Beckman JA-12 rotor at 25^{o}C. Each pellet was then washed 3 times by resuspending in 50 ml of PBS (pH 7.2) and centrifuging as

described above. After the final wash, the cell suspensions were heated to 85^{o} C and then filtered through 0.45 μm syringe filters (MILLEX HV, Millipore, Carrigtwohill, Co. Cork, Ireland) in 25 ml aliquots. The filtrate was collected in fresh conical centrifuge tubes which were then centrifuged at 6000 rpm for 30 min in a Beckman JA-12 rotor at 4^{o} C. Small white pellets were resuspended in a total volume of 1 ml of PBS (pH 7.2), placed in an Eppendorf tube, and centrifuged at 13,000 rpm in a benchtop microcentrifuge. The single cell pellet was resuspended in either PBS (pH 7.2) for microscopic analysis or DNA isolation.

Genome Sequencing and Assembly. Library construction, high-throughput sequencing, and genome assembly were performed at the Joint Genome Institute, Walnut Creek, CA, USA. Fosmid libraries were constructed from the raw OPF1 enrichment culture DNA and DNA from filtered cells using the pCC1FOS based system (Epicentre Biotechnologies, Madison, WI, USA). Small-insert libraries (2-3 kb insert size) were produced for raw, filtered, and purified cell samples of "*Ca.* K. cryptofilum". DNA sequencing was performed with BigDye Terminators v3.1 and resolved with ABI PRISM 3730 DNA sequencers (PE Applied Biosystems, Foster City, CA, USA). The sequencing reads were base called using phred version 0.990722.g (18, 19), vector trimmed using crossmatch SPS-3.57, and assembled into contigs using parallel phrap (www.phrap.org). The contigs were aligned with an 84.7 kb "*Ca.* K. cryptofilum" genomic sequence containing the pOPF_08 16S rRNA gene, and the 16S rRNA genes from *Thermofilum pendens* and *Thermosphaera aggregans*. The resulting assembly was binned by scaffolding and read depth, and by aligning binned scaffolds against the 16S rRNA sequences from the predominant archaeal genomes in the mixture and an 84.7 kb reference contig.

Gene Prediction and Comparative Analysis. An automated annotation was provided by the Computational Biology Group at Oak Ridge National Laboratory (ORNL), Oak Ridge, TN, USA. An additional analysis was performed by loading the complete korarchaeal genome into the Integrated Microbial Genomes (IMG) analysis tool developed by the Joint Genome Institute, Walnut Creek, CA, USA (20). To identify putative functions for the predicted open reading frames, the gene models were compared against all non-redundant (NR) sequences deposited in the National Center for Biotechnology Information (NCBI) database using BLASTP (21). The gene models were also compared against the protein families (Pfam) (22), Clusters of Orthologous Groups (COGs) (23), and the Kyoto Encyclopedia of Genes and Genomes (KEGG) (24) databases for further functional information. The BLAST E-value threshold was set at 1e-05. The Pfam threshold was also set at 1e-05. The COG search was performed using the default cutoff. For tRNA identification, tRNAscan-SE was run using the prokaryotic default settings (25), followed by manual analyses. DNA and amino acid sequence alignments were generated with CLUSTALW using the default parameters (26) and checked and edited manually using BioEdit v7.0.5 (http://www.mbio.ncsu.edu/BioEdit/bioedit.html).

Phylogenetic Analyses. An alignment of concatenated small and large subunit rRNA sequences (SSU+LSU rRNA) was constructed based on their conserved secondary structures and refined by hand using BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Highly variable loop regions that could not be confidently aligned as well as regions in which some sequences were incomplete were masked out. The final alignment contained 27 taxa and 2965 nucleotide characters and was analyzed using the software Modeltest v3.7 to determine the best evolutionary

model that fits the data (27). The likelihood ratio test statistics indicated that to correspond to the general time reversible model with unequal nucleotide frequencies, a fraction of invariable sites and with six independent substitution rates that follow a gamma distribution (GTR+I+G). Phylogenetic analysis was performed using PAUP* 4.0b8 (Sinauer Associates, Sunderland, MA). The initial GTR+I+G parameters were determined using five rounds of neighbor-joining tree construction/parameter estimation and were used for an unrestricted heuristic maximum likelihood tree search with random sequence addition and tree bisection-reconnection (TBR) branch swapping. The GTR+I+G parameters were estimated from the resulting tree and used for a new round of heuristic search that used five independent rounds of random sequence additions and TBR branch swapping, resulting in a single maximum likelihood tree. Bootstrapping (100 replicates) was conducted using the same parameters and one round or random sequence addition/TBR branch swapping for each replicate.

Protein sequences were aligned on the M-Coffee web server using a combination of multiple sequence alignment methods (28). The alignments were visually inspected and the regions not confidently aligned or containing high variability were masked out. The final alignments contained 476 aminoacid positions for EF2 and 211 aminoacid positions for FtsZ genes, respectively. Rooted (Fig. 2*B*) and unrooted (SI Fig. 8) EF2 trees were both calculated. Phylogenetic reconstruction was performed using two different approaches. First, the amino acid substitution model was chosen with the software Modelgenerator v84, which uses Akaike and Bayesian information criteria (AIC and BIC) to select the optimum among 88 different models (29). For all the datasets the optimum model was a combination of the RtRev substitution matrix with an estimated fraction of invariable sites and with six categories of substitution rates following a gamma distribution with the shape parameter optimized based on the data. These parameters were used to calculate maximum likelihood trees and non-parametric bootstrap support for the individual nodes with the PhyML v2.4.4 software (30). Since accurate reconstruction of deep phylogenetic branching order is complicated due to long branch attraction artifacts, we also used tree inference with PhyloBayes v2.3 (31). That software employs a Bayesian Monte Carlo Markov Chain (MCMC) sampler and a site-heterogeneous mixture model (CAT) and has been shown to be less prone to LBA (31, 32). Two to four independent chains were run in parallel for at least 10,000 generations starting from random trees, until convergence (maximum discrepancy across all bipartions of <0.05). The first 5,000 cycles were then discarded as the burn-in and the remaining ones were sampled each of every ten to calculate majority-rule posterior consensus trees with bipartition frequencies at each node.

Concatenated r-protein and RNAP trees: Using the COGNITOR method (33, 34), representatives of three bacterial species (*E. coli*, *B. subtilis* and *D. radiodurans*) were added to the arCOGs that include 33 universally conserved ribosomal proteins and three largest RNA polymerase subunits. The sequences were aligned using MUSCLE (35). Start codon positions were verified and, when necessary, corrected manually. Maximum Likelihood trees for the concatenated alignment of 36 proteins (8057 aminoacid positions, 5696 for ribosomal proteins and 2361 for RNA polymerase subunits) were constructed using TreeFinder (14), with the Whelan and Goldman (WAG) evolutionary model and gamma-distributed site rates. Tree topologies were compared using the Approximately Unbiased test (AU) (13) implemented in the TreeFinder program. The topology

where "*Ca.* K. cryptofilum" is a sister group of all Crenarchaeota, including *Cenarchaeum symbiosum* passed the 0.05 p-value cut-off (AU p-value = 0.17).

Supplementary References:

1. Burggraf S, Huber H, Stetter KO (1997) Reclassification of the crenarchael orders and families in accordance with 16S rRNA sequence data. *Int J Syst Bacteriol* 47:657-660.
2. Spear JR, Walker JJ, McCollom TM, Pace NR (2005) Hydrogen and bioenergetics in the Yellowstone geothermal ecosystem. *Proc Natl Acad Sci U S A* 102:2555-2560.
3. Barns SM, Fundyga RE, Jeffries MW, Pace NR (1994) Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc Natl Acad Sci U S A* 91:1609-1613.
4. Reysenbach AL, Ehringer M, Hershberger K (2000) Microbial diversity at 83 degrees C in Calcite Springs, Yellowstone National Park: another environment where the Aquificales and "Korarchaeota" coexist. *Extremophiles* 4:61-67.
5. Raskin L, Stromley JM, Rittmann BE, Stahl DA (1994) Group-specific 16S rRNA hybridization probes to describe natural communities of methanogens. *Appl Environ Microbiol* 60:1232-1240.
6. Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* 56:1919-1925.
7. Nishida H, Nishiyama M, Kobashi N, Kosuge T, Hoshino T, Yamane H (1999) A prokaryotic gene cluster involved in synthesis of lysine through the amino adipate pathway: a key to the evolution of amino acid biosynthesis. *Genome Research* 9:1175-1183.
8. White RH (2004) L-Aspartate semialdehyde and a 6-deoxy-5-ketohexose 1-phosphate are the precursors to the aromatic amino acids in *Methanocaldococcus jannaschii*. *Biochemistry* 43:7618-7627.
9. Mahillon J, Chandler M (1998) Insertion sequences. *Microbiol Mol Biol Rev* 62:725-774.
10. Filee J, Siguier P, Chandler M (2007) Insertion sequence diversity in archaea. *Microbiol Mol Biol Rev* 71:121-157.
11. She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ, Chan-Weiher CC, Clausen IG, Curtis BA, De Moors A, Erauso G, Fletcher C, Gordon PM, Heikamp-de Jong I, Jeffries AC, Kozera CJ, Medina N, Peng X, Thi-Ngoc HP, Redder P, Schenk ME, Theriault C, Tolstrup N, Charlebois RL, Doolittle WF, Duguet M, Gaasterland T, Garrett RA, Ragan MA, Sensen CW, Van der Oost J (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci U S A* 98:7835-7840.
12. Greve B, Jensen S, Brugger K, Zillig W, Garrett RA (2004) Genomic comparison of archaeal conjugative plasmids from Sulfolobus. *Archaea* 1:231-239.
13. Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. *Syst Biol* 51:492-508.
14. Jobb G, von Haeseler A, Strimmer K (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* 4:18.
15. Brochier C, Forterre P, Gribaldo S (2004) Archaeal phylogeny based on proteins of the transcription and translation machineries: tackling the *Methanopyrus kandleri* paradox. *Genome Biol* 5:R17.

16. Shock EL, Holland M, Meyer-Dombard D, Amend JP (2005) in *Geothermal Biology and Geochemstry in Yellowstone National Park*, ed. Inskeep WP (Thermal Biology Institute, Montana State University, Bozeman), pp. 95-112.
17. Allen MB (1959) Studies with Cyanidium caldarium, an anomalously pigmented chlorophyte. *Arch Mikrobiol* 32:270-277.
18. Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 8:186-194.
19. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175-185.
20. Markowitz VM, Korzeniewski F, Palaniappan K, Szeto E, Werner G, Padki A, Zhao X, Dubchak I, Hugenholtz P, Anderson I, Lykidis A, Mavromatis K, Ivanova N, Kyrpides NC (2006) The integrated microbial genomes (IMG) system. *Nucleic Acids Res* 34:D344-348.
21. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389-3402.
22. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A (2006) Pfam: clans, web tools and services. *Nucleic Acids Res* 34:D247-251.
23. Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 28:33-36.
24. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* 34:D354-357.
25. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955-964.
26. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31:3497-3500.
27. Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817-818.
28. Moretti S, Armougom F, Wallace IM, Higgins DG, Jongeneel CV, Notredame C (2007) The M-Coffee web server: a meta-method for computing multiple sequence alignments by combining alternative alignment methods. *Nucleic Acids Res* 35:W645-648.
29. Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McLnerney JO (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol* 6:29.
30. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696-704.
31. Lartillot N, Brinkmann H, Philippe H (2007) Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol* 7 Suppl 1:S4.
32. Philippe H, Brinkmann H, Martinez P, Riutort M, Baguna J (2007) Acoel flatworms are not platyhelminthes: evidence from phylogenomics. *PLoS ONE* 2:e717.
33. Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. *Science* 278:631-637.

34. Makarova KS, Sorokin AV, Novichkov PS, Wolf YI, Koonin EV (2007) Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea. *Biol Direct* 2:33.
35. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792-1797.
36. Brunk CF, Eis N (1998) Quantitative measure of small-subunit rRNA gene sequences of the kingdom korarchaeota. *Appl Environ Microbiol* 64:5064-5066.
37. Baker GC, Smith JJ, Cowan DA (2003) Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* 55:541-555.

Supporting Figure Legends:

Fig. 3. Neighbour-Joining tree constructed from distance matrix analysis of SSU rDNA sequences obtained from the Obsidian Pool enrichment culture. "*Ca.* K. cryptofilum" is represented by clone pOPF_08.

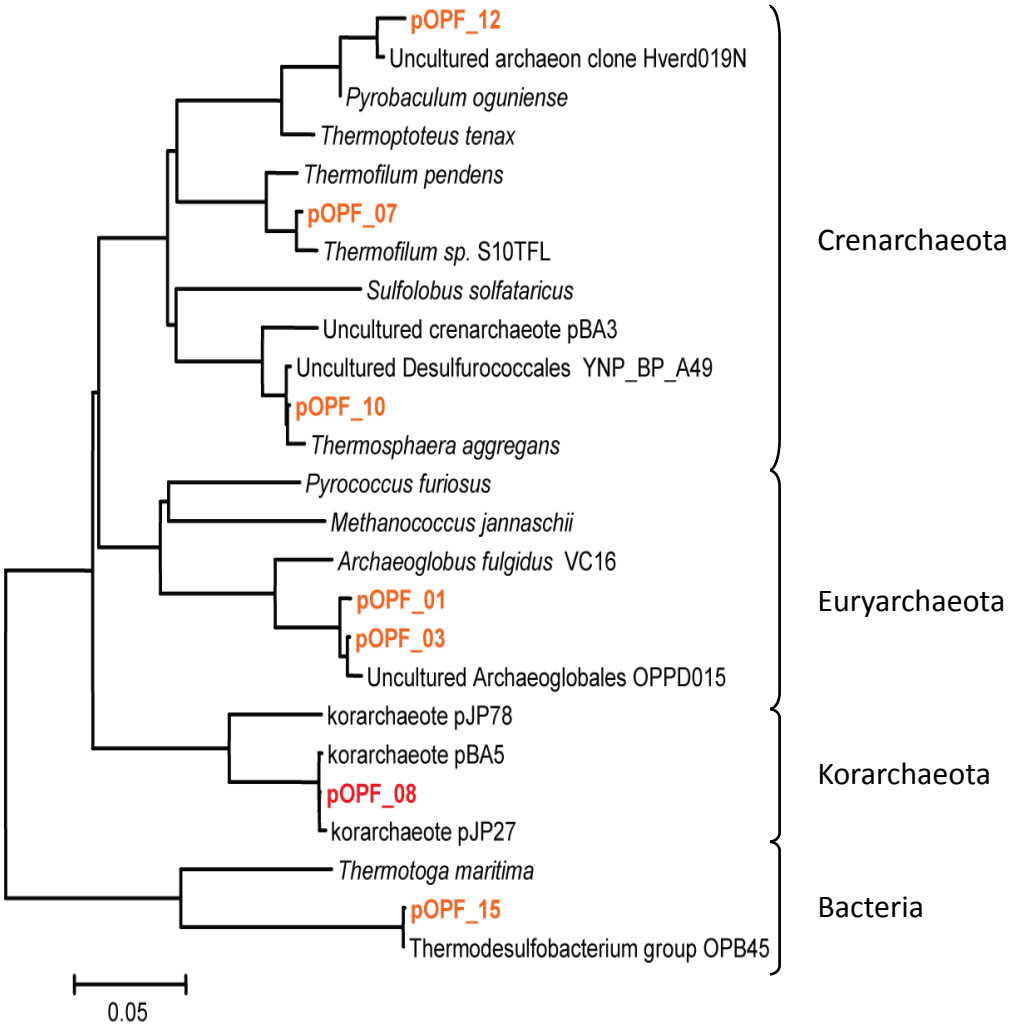
Fig. 4. Fluorescence *in-situ* hybridization analysis using crenarchaeal and *Korarchaeota* specific probes. (*A*) Pure culture of *Thermofilum* sp. S10TFL hybridizing to Alexa488-labeled CREN499R (CCARNCTTGCCCCCCGCT) probe. (*B*) CREN499R probe applied to Obsidian Pool enrichment culture. (*C*) Cy3-labeled *Korarchaeota* specific probes KR515R (CCAGCCTTACCCTCCCCT) and KR565R (AGTATGCGTGGGAACCCCTC) hybridizing to cells of "*Ca.* K. cryptofilum". (*D*) Phase contrast image of cell preparation. Scale bars represent 5µm.

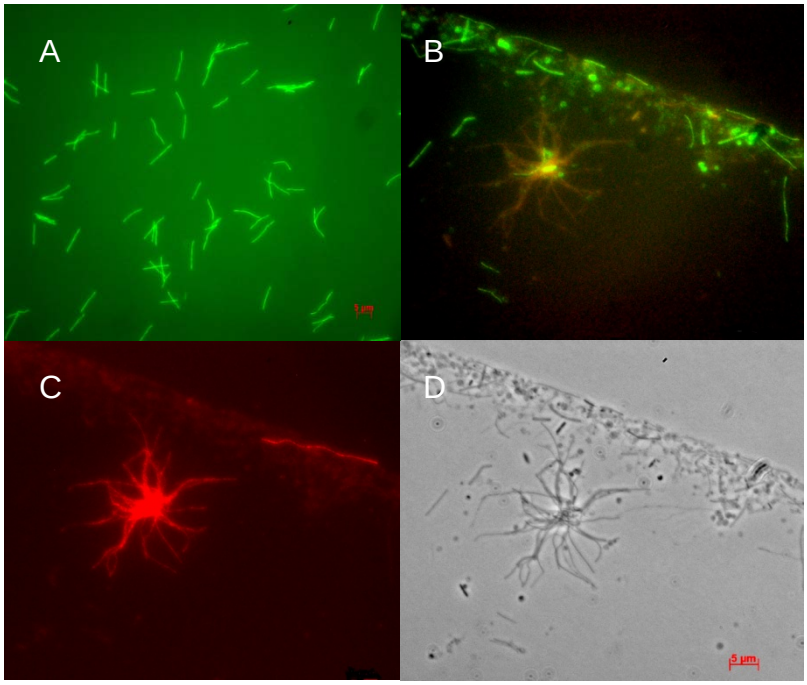
Fig. 5. Relative abundance of PCR amplified SSU rDNAs obtained from (*A*), untreated, Obsidian Pool enrichment culture and (*B*), filtered cells exposed to 0.2% SDS for 15 min. PCR primers included Ab779F (GCRAASSGGATTAGATACCC) (36) and UA1406R (ACGGGCGGTGWGTRCAA) (37). The treatment resulted in a highly enriched cell fraction of cells represented by the pOPF_08 SSU rDNA sequence type.

Fig. 6. Prior to WGS sequencing, fosmid clones containing the pOPF_08 SSU rRNA sequence were retrieved and assembled from a library constructed from Obsidian Pool enrichment culture DNA with a CopyControl fosmid library kit (Epicentre, Madison, WI, USA). (*A*) Fosmid clone pKOR01M18 containing korarchaeal small and large subunit rRNA genes was retrieved by probing the library with a 209 bp, 5' region of the pOPF_08 sequence. The LSU rRNA gene contains a self splicing, intron- encoded LAGLIDADG type endonuclease. (*B*) Additional clones were retrieved by probing the library with labelled end-sequences from the finished fosmids. Overlapping fosmid clones were assembled to construct an 84.7 kb genomic contig.

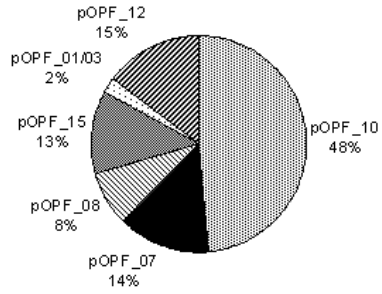
Fig. 7. Majority-rule posterior consensus tree (PhyloBayes) for FtsZ protein sequences. Values at the nodes indicate the posterior probability for that branching (the branch was collapsed when the value was <50). Scale bar indicates the inferred number of substitutions per site.

Fig. 8. Unrooted archaeal phylogeny based on translation elongation factor 2 (EF2) proteins. The numbers indicate bootstrap support for PhyML/consensus posterior probability (Phyloblast), an asterisk indicates <50 bootstrap support. Where both values were <50, the branch was collapsed. Scale bar indicates the inferred number of substitutions per site.


pOPF_12
Uncultured archaeon clone Hverd019N
Pyrobaculum oguniense
Thermoptoteus tenax
Thermofilum pendens
pOPF_07
Thermofilum sp. S10TFL
Sulfolobus solfataricus
Uncultured crenarchaeote pBA3
Uncultured Desulfurococcales YNP_BP_A49
pOPF_10
Thermosphaera aggregans
Pyrococcus furiosus
Methanococcus jannaschii
Archaeoglobus fulgidus VC16
pOPF_01
pOPF_03
Uncultured Archaeoglobales OPPD015
korarchaeote pJP78
korarchaeote pBA5
pOPF_08
korarchaeote pJP27
Thermotoga maritima
pOPF_15
Thermodesulfobacterium group OPB45
0.05
Crenarchaeota
Euryarchaeota
Korarchaeota
Bacteria

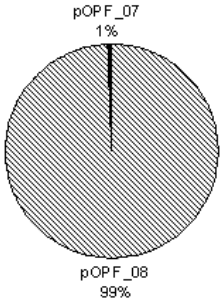

A
B
C
D

A.


pOPF_12
15%
pOPF_01/03
2%
pOPF_15
13%
pOPF_08
8%
pOPF_07
14%
pOPF_10
48%

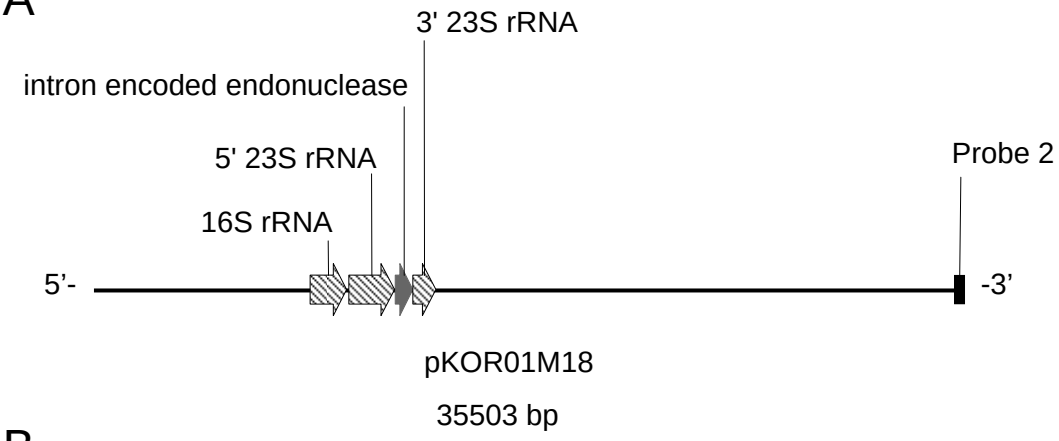
N=176

B.

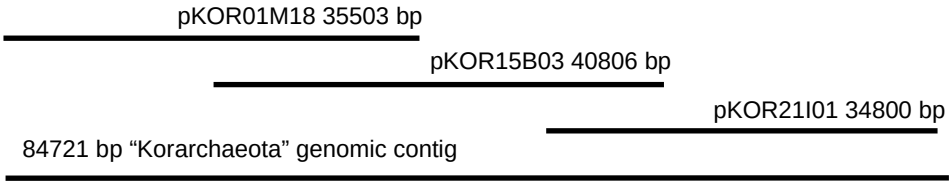

pOPF_07
1%
pOPF_08
99%

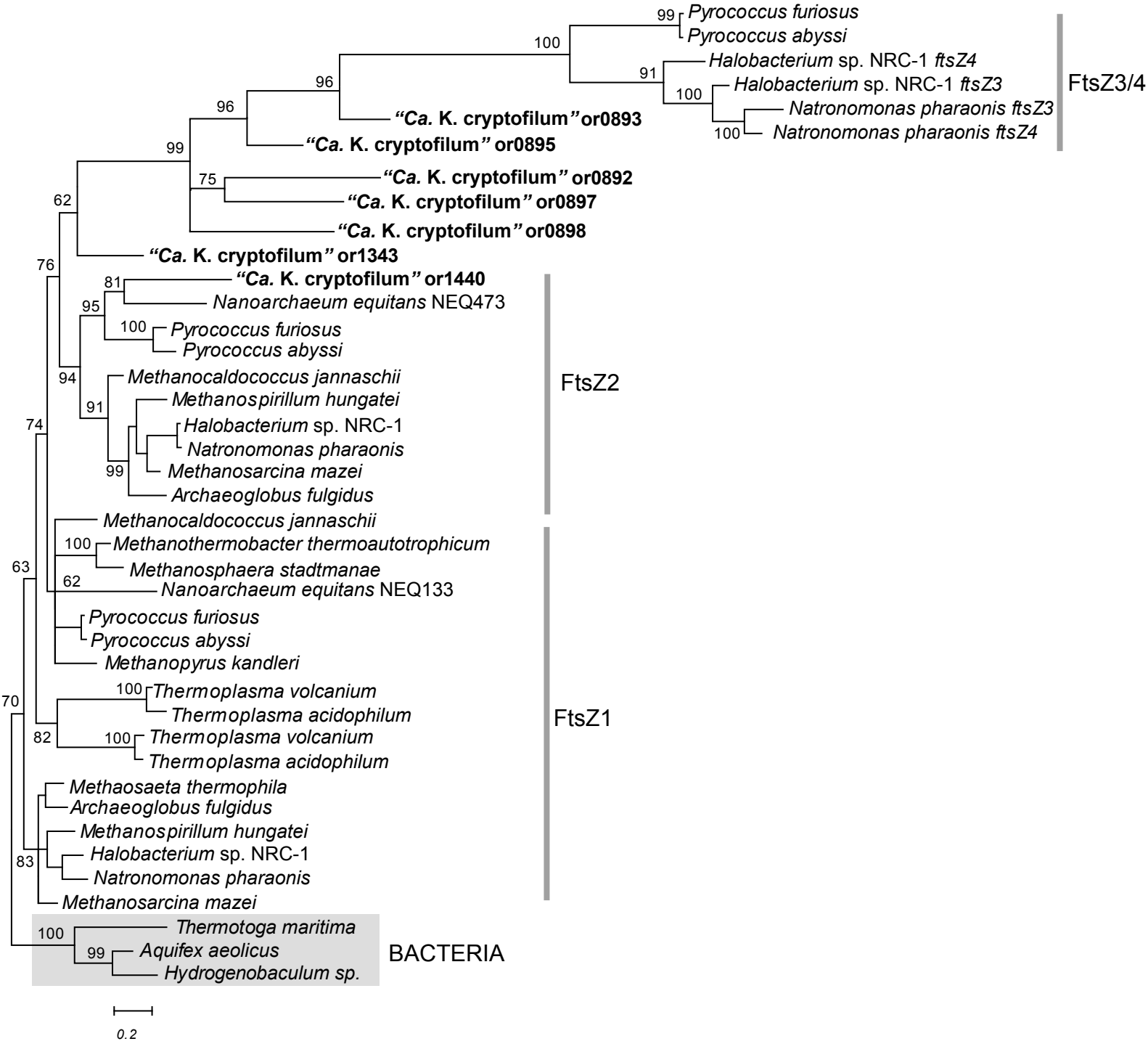
N=180

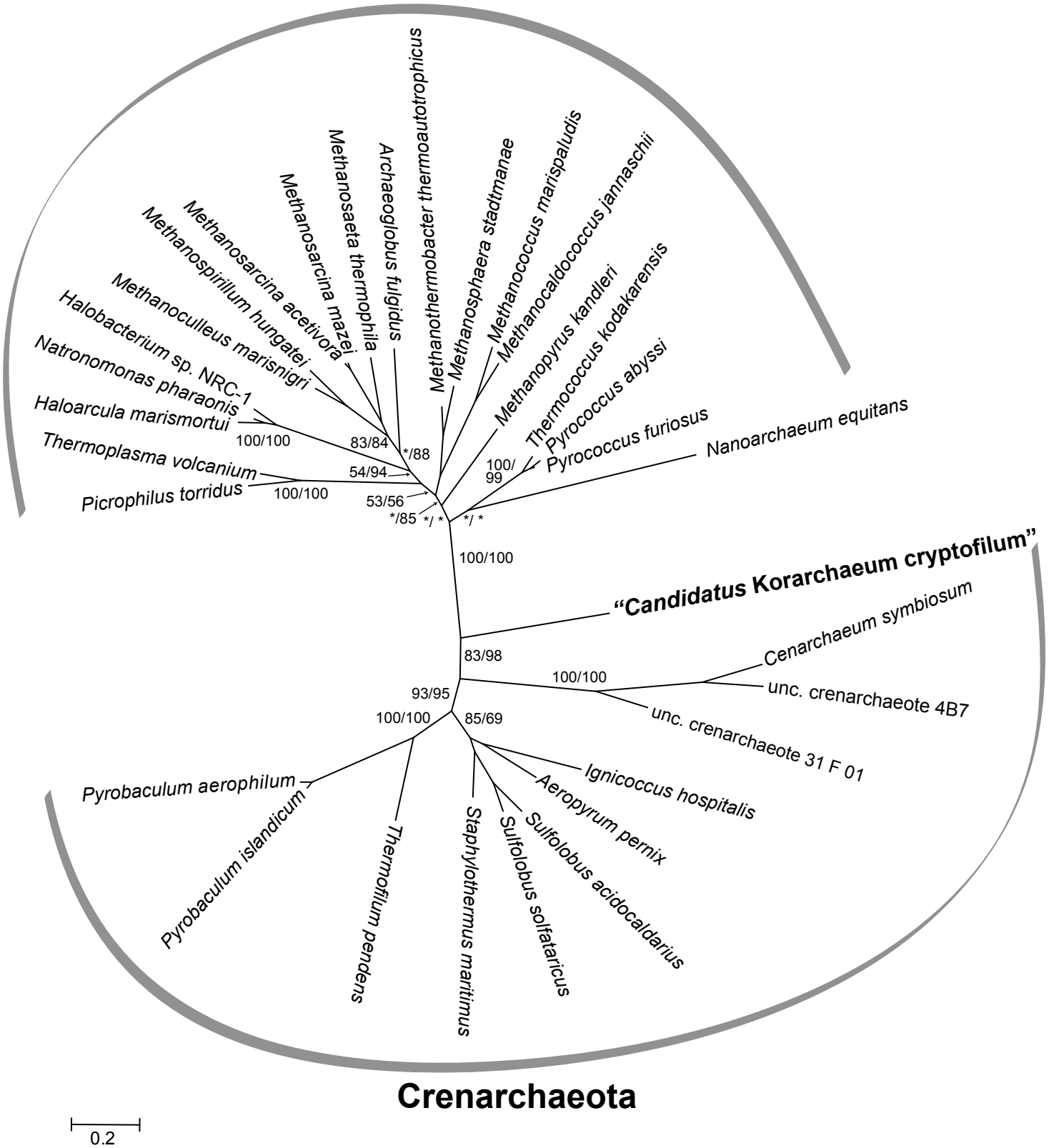
A


3' 23S rRNA
intron encoded endonuclease
5' 23S rRNA
Probe 2
16S rRNA
5'-
-3'
pKOR01M18
35503 bp

B


pKOR01M18 35503 bp
pKOR15B03 40806 bp
pKOR21I01 34800 bp
84721 bp “Korarchaeota” genomic contig


99
Pyrococcus furiosus
Pyrococcus abyssi
100
Halobacterium sp. NRC-1 ftsZ4
91
96
Halobacterium sp. NRC-1 ftsZ3
100
96
Natronomonas pharaonis ftsZ3
100
Natronomonas pharaonis ftsZ4
"Ca. K. cryptofilum" or0893
"Ca. K. cryptofilum" or0895
99
75
"Ca. K. cryptofilum" or0892
62
"Ca. K. cryptofilum" or0897
"Ca. K. cryptofilum" or0898
"Ca. K. cryptofilum" or1343
76
81
"Ca. K. cryptofilum" or1440
Nanoarchaeum equitans NEQ473
95
100
Pyrococcus furiosus
Pyrococcus abyssi
94
Methanocaldococcus jannaschii
Methanospirillum hungatei
91
74
Halobacterium sp. NRC-1
Natronomonas pharaonis
99
Methanosarcina mazei
Archaeoglobus fulgidus
Methanocaldococcus jannaschii
100
Methanothermobacter thermoautotrophicum
63
Methanosphaera stadtmanae
62
Nanoarchaeum equitans NEQ133
Pyrococcus furiosus
Pyrococcus abyssi
Methanopyrus kandleri
100
Thermoplasma volcanium
70
Thermoplasma acidophilum
82
100
Thermoplasma volcanium
Thermoplasma acidophilum
Methaosaeta thermophila
Archaeoglobus fulgidus
Methanospirillum hungatei
83
Halobacterium sp. NRC-1
Natronomonas pharaonis
Methanosarcina mazei
100
Thermotoga maritima
99
Aquifex aeolicus
Hydrogenobaculum sp.
0.2
FtsZ3/4
FtsZ2
FtsZ1
BACTERIA


Euryarchaeota
Methanothermobacter thermoautotrophicus
Archaeoglobus fulgidus
Methanosaeta thermophila
Methanosarcina mazei
Methanosarcina acetivora
Methanospirillum hungatei
Methanoculleus marisnigri
Halobacterium sp. NRC-1
Natronomonas pharaonis
Haloarcula marismortui
Thermoplasma volcanium
Picrophilus torridus
Methanosphaera stadtmanae
Methanococcus marispaludis
Methanocaldococcus jannaschii
Methanopyrus kandleri
Thermococcus kodakarensis
Pyrococcus abyssi
Pyrococcus furiosus
Nanoarchaeum equitans
100/100
83/84
*/88
54/94
100/100
53/56
*/85
*/ *
100/
99
*/ *
100/100
"Candidatus Korarchaeum cryptofilum"
83/98
Cenarchaeum symbiosum
100/100
unc. crenarchaeote 4B7
93/95
unc. crenarchaeote 31 F 01
100/100
85/69
Ignicoccus hospitalis
Aeropyrum pernix
Sulfolobus acidocaldarius
Sulfolobus solfataricus
Staphylothermus maritimus
Thermofilum pendens
Pyrobaculum islandicum
Pyrobaculum aerophilum
Crenarchaeota
0.2

Gene	Strand	Start	Stop	Size (a.a.)	Product	Gene Name	COGs	arCOGs	COG_gene	Func	arCOG annotation	# of Cren in arCOGs	#
Kcr_0001	+	127	1161	345	TGS			arCOG00358	DRG	R	Predicted GTPase	13	2
Kcr_0002	+	1162	1839	226	ribose 5-phosphate isomerase	rpiA		arCOG01122	RpiA	G	Ribose 5-phosphate isomerase	13	2
Kcr_0003	-	2202	2429	76	4Fe-4S ferredoxin, iron-sulfur binding			arCOG00349	Fer	C	Ferredoxin	8	1
Kcr_0004	-	2519	2971	151	iron dependent repressor			arCOG02100	TroR	K	Mn-dependent transcriptional regulator (DtxR family)	10	2
Kcr_0005	+	3026	3535	170	molybdopterin binding domain			arCOG00214	MoaB	H	Molybdopterin biosynthesis enzyme	12	2
Kcr_0006	-	3525	5477	651	MoeA-like, domain I and II			arCOG00217	MoeA	H	Molybdopterin biosynthesis enzyme	10	2
Kcr_0007	-	5452	6714	421	MoeA-like, domain I and II			arCOG00216	MoeA	H	Molybdopterin biosynthesis enzyme	12	2
Kcr_0008	+	6797	7813	339	Homoserine dehydrogenase			arCOG01351	ThrA	E	Homoserine dehydrogenase	10	2
Kcr_0009	+	7813	9165	451	aspartate kinase	aspK		arCOG00861	LysC	E	Aspartokinase	11	2
Kcr_0010	-	9324	10079	252	inositol monophosphatase			arCOG01349	SuhB	G	Archaeal fructose-1,6-bisphosphatase or related enzyme of inositol monophosphatase family	13	2
Kcr_0011	-	10089	10436	116	ATP-cone			arCOG03717/1	-	R	ATP-cone domain	3	1
Kcr_0012	-	10803	11090	96	thiamineS			arCOG00536	MoaD	H	Molybdopterin converting factor, small subunit	9	2
Kcr_0013	-	11099	12961	621	aldehyde ferredoxin oxidoreductase			arCOG00707	-	C	Aldehyde:ferredoxin oxidoreductase	8	7
Kcr_0014	+	13022	13939	306	Radical SAM			arCOG00662	-	R	Biotin synthase-related enzyme	5	1
Kcr_0015	+	14022	14732	237	Radical SAM			arCOG05825	-	R	Radical SAM superfamily enzyme	5	4
Kcr_0016	-	14725	16392	556	Radical SAM			arCOG01364	-	C	Radical SAM superfamily enzyme	3	2
Kcr_0017	+	17140	17517	126	Forkhead-associated			arCOG01694/1	-	T	FHA domain	3	2
Kcr_0018	+	17514	18617	368	protein kinase			arCOG03683/3	SPS1	R	Serine/threonine protein kinase	1	9
Kcr_0019	+	18667	19461	265	Cobyrinic acid a,c-diamide synthase			arCOG00585	Mrp	D	Mrp family protein, ATPase	13	2
Kcr_0020	+	19723	21264	514	PKD			arCOG02508/3	-	S	PKD repeat	0	7
Kcr_0021	-	21266	21532	89	transcriptional regulator TrmB			arCOG04056	ArsR	K	Transcriptional regulator containing HTH domain, ArsR family	7	0
Kcr_0022	-	21542	22606	355	Radical SAM			arCOG00938	-	R	Radical SAM superfamily enzyme	13	2
Kcr_0023	+	22683	23891	403	Methionine adenosyltransferase			arCOG01678	MetK	E	Archaeal S-adenosylmethionine synthetase	12	2
Kcr_0024	-	23863	26946	1028	DEAD/DEAH box helicase-like			arCOG00555/3	-	R	Distinct helicase family with a unique C-terminal domain including a metal-binding cysteine cluster	12	2
Kcr_0025	-	26952	27380	143	conserved hypothetical protein			arCOG01832	IbpA	O	Molecular chaperone (HSP20 family)	11	2
Kcr_0026	-	27415	30123	903	SMC protein-like			arCOG00368	SbcC	L	ATPase involved in DNA repair, SbcC	13	2
Kcr_0027	-	30120	31289	390	metallophosphoesterase			arCOG00397	SbcD	L	DNA repair exonuclease, SbcD	13	2
Kcr_0028	-	31291	32883	531	protein of unknown function DUF87			arCOG00280	-	L	HerA helicase	13	2
Kcr_0029	-	32873	33919	349	hypothetical protein			arCOG00367	-	L	NurA 5'-3' nuclease	13	2
Kcr_0030	+	34023	35402	460	CoA-binding			arCOG01340	-	C	Acyl-CoA synthetase (NDP forming)	11	2
Kcr_0031	+	35513	36220	236	PspA/IM30			arCOG04782	PspA	K	Phage shock protein A	0	9
Kcr_0032	+	36222	36623	134	hypothetical protein								
Kcr_0033	-	36618	36944	109	t-RNA-binding region			arCOG01136	ARC1	R	EMAP domain RNA-binding protein	12	8
Kcr_0034	+	37032	38306	425	beta-lactamase-like			arCOG00541	YSH1	J	Predicted exonuclease of the beta-lactamase fold involved in RNA processing	13	1
Kcr_0035	-	38303	39373	357	Radical SAM			arCOG00932	-	R	Uncharacterized conserved protein related to pyruvate formate-lyase activating enzyme	10	2
Kcr_0036	-	39375	40613	413	adenosylhomocysteinase	ahcY		arCOG04137	SAM1	H	S-adenosylhomocysteine hydrolase	13	2
Kcr_0037	-	40616	40840	75	hypothetical protein			arCOG02062	SirA	O	Predicted redox protein, regulator of disulfide bond formation	9	1
Kcr_0038	-	40837	41385	183	ribosomal-protein-alanine acetyltransferase	rimI		arCOG00833	RimI	R	Acetyltransferase (GNAT) family	13	2
Kcr_0039	+	41488	42288	267	hypothetical protein								
Kcr_0040	+	42321	42578	86	similarity to 40S RIBOSOMAL PROTEIN S25			arCOG04327/2	-	J	Ribosomal protein S25	13	0
Kcr_0041	+	42611	42997	129	hypothetical protein			arCOG04488	-	S	Uncharacterized conserved protein	1	3
Kcr_0042	+	42994	43632	213	beta-lactamase-like			arCOG00497	-	R	Zn-dependent hydrolase of the beta-lactamase fold	12	2
Kcr_0043	-	43616	44578	321	thiamine-monophosphate kinase	thiL		arCOG00638	ThiL	H	Thiamine monophosphate kinase	13	2
Kcr_0044	+	44636	45958	441	conserved hypothetical protein			arCOG01229	-	R	Predicted GTPase	8	5
Kcr_0045	+	45955	46701	249	Creatininase			arCOG04536	-	R	Uncharacterized protein, putative amidase	13	1
Kcr_0046	-	46696	47421	242	hypothetical protein			arCOG01897/3	IolE	G	Sugar phosphate isomerase/epimerase	0	7
Kcr_0047	+	48192	48422	77	transcription factor CBF/NF-Y histone			arCOG02144	HHT1	L	Histones H3 and H4	3	2
Kcr_0048	+	48817	49665	283	Xylose isomerase-like TIM barrel			arCOG01895	IolE	G	Sugar phosphate isomerase/epimerase	4	2
Kcr_0049	-	49609	50172	188	PAP2 family protein		COG0671			I			
Kcr_0050	+	50228	51058	277	Protein of unknown function DUF516			arCOG01616	-	S	Uncharacterized protein conserved in archaea	12	2
Kcr_0051	+	51045	51383	113	hypothetical protein								
Kcr_0052	+	51650	51955	102	protein of unknown function DUF211			arCOG04140	-	S	Uncharacterized protein conserved in archaea	11	1
Kcr_0053	+	51952	53250	433	MiaB-like tRNA modifying enzyme			arCOG01358	MiaB	J	2-methylthioadenine synthetase	13	2
Kcr_0054	-	53243	53611	123	S-adenosylmethionine decarboxylase related			arCOG00279	SpeD	E	S-adenosylmethionine decarboxylase	12	1
Kcr_0055	-	53601	54269	223	beta-lactamase-like			arCOG00497	-	R	Zn-dependent hydrolase of the beta-lactamase fold	12	2
Kcr_0056	-	54319	54957	213	hypothetical protein								
Kcr_0057	+	55015	56736	574	Radical SAM			arCOG00933	-	R	Predicted Fe-S oxidoreductase	13	1
Kcr_0058	+	56737	57888	384	geranylgeranyl reductase			arCOG00570	FixC	C	Dehydrogenase (flavoprotein)	13	2
Kcr_0059	-	57815	58846	344	hypothetical protein								
Kcr_0060	-	58851	60455	535	Phosphoenolpyruvate synthase/pyruvate phosphate dikinase-like			arCOG01111/3	PpsA	G	Phosphoenolpyruvate synthase/pyruvate phosphate dikinase	12	2
Kcr_0061	-	60452	61225	258	protein of unknown function DUF62			arCOG04309	-	S	Uncharacterized conserved protein	12	1
Kcr_0062	-	61482	62045	188	2',5' RNA ligase			arCOG01736	LigT	J	2'-5' RNA ligase	13	2
Kcr_0063	-	62050	62403	118	Cupin 2, conserved barrel			arCOG02994	-	S	Cupin domain containing protein	5	1
Kcr_0064	-	62523	63410	296	phosphoesterase, DHHA1			arCOG00431/3	-	R	Predicted phosphohydrolase (DHH superfamily)	6	0
Kcr_0065	+	63463	63906	148	Cytidyltransferase-related			arCOG01222	TagD	M	Cytidylyltransferase	13	2
Kcr_0066	+	64408	66087	560	methionyl-tRNA synthetase	metG		arCOG00810	MetG	J	Methionyl-tRNA synthetase	13	2
Kcr_0067	-	66084	66782	233	hypothetical protein								
Kcr_0068	+	67015	67959	315	Glycerol-1-phosphate dehydrogenase [NAD(P)]			arCOG00982	GldA	C	Glycerol dehydrogenase or related enzyme	13	2
Kcr_0069	-	67940	68554	205	hypothetical protein								
Kcr_0070	-	68605	70380	592	ABC transporter related			arCOG00187	-	R	Predicted ATPase, RNase L inhibitor (RLI) homolog	13	2
Kcr_0071	-	70385	70987	201	pyruvate ferredoxin/flavodoxin oxidoreductase			arCOG01602	PorG	C	Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidoreductase, gamma subunit	11	2
Kcr_0072	-	70990	72939	650	thiamine pyrophosphate enzyme-like TPP-binding			arCOG01609	-	C	Indolepyruvate ferredoxin oxidoreductase, alpha and beta subunit	11	2
Kcr_0073	-	73051	74346	432	CoA-binding			arCOG01340	-	C	Acyl-CoA synthetase (NDP forming)	11	2
Kcr_0074	-	74353	74871	173	Uncharacterized domain 2			arCOG00985	-	J	Predicted RNA-binding protein (contains PUA domain)	13	2
Kcr_0075	-	74986	75831	282	Xylose isomerase-like TIM barrel			arCOG01894	Nfo	L	Endonuclease IV	13	2
Kcr_0076	+	76209	77273	355	Radical SAM			arCOG00934	PflX	C	Radical SAM superfamily enzyme	9	1
Kcr_0077	-	77263	77832	190	adenylate kinase related protein			arCOG01045	CoaE	H	Dephospho-CoA kinase	13	2
Kcr_0078	+	77905	78720	272	CRISPR-associated protein, TM1814			arCOG01439	-	V	CRISPR system related protein, RAMP superfamily	8	0
Kcr_0079	-	78724	79116	131	S-adenosylmethionine decarboxylase related			arCOG00279	SpeD	E	S-adenosylmethionine decarboxylase	12	1
Kcr_0080	-	79521	80558	346	hypothetical protein								
Kcr_0081	-	80771	82051	427	aminotransferase class-III			arCOG00915	GabT	E	4-aminobutyrate aminotransferase or related aminotransferase	12	1
Kcr_0082	-	82117	82839	241	ABC transporter related			arCOG00924	LivF	E	ABC-type branched-chain amino acid transport system, ATPase component	9	6
Kcr_0083	-	82844	83683	280	ABC transporter related			arCOG00925	LivG	E	ABC-type branched-chain amino acid transport system, ATPase component	9	6
Kcr_0084	-	83772	85187	472	Extracellular ligand-binding receptor			arCOG01021	LivK	E	ABC-type branched-chain amino acid transport system, periplasmic component	9	8
Kcr_0085	+	85236	86312	359	inner-membrane translocator			arCOG01269	LivH	E	Branched-chain amino acid ABC-type transport system, permease component	8	6
Kcr_0086	+	86318	87409	364	inner-membrane translocator			arCOG01273	LivM	E	ABC-type branched-chain amino acid transport system, permease component	8	5
Kcr_0087	-	87421	88689	423	Glu/Leu/Phe/Val dehydrogenase, C terminal			arCOG01352	GdhA	E	Glutamate dehydrogenase/leucine dehydrogenase	13	1
Kcr_0088	+	88845	89273	143	hypothetical protein								
Kcr_0089	+	89270	90259	330	hypothetical protein			arCOG06251	FtsI,FolC	M	DP-N-acetylmuramyl tripeptide synthase and Folylpolyglutamate synthase	0	2
Kcr_0090	-	90246	91679	478	Radical SAM			arCOG00940	-	R	Radical SAM superfamily enzyme	2	7
Kcr_0091	+	91941	92369	143	nucleic acid binding, OB-fold, tRNA/helicase-type			arCOG01510/2	RFA1	L	Single-stranded DNA-binding replication protein A (RPA), large (70 kD) subunit or related ssDNA-binding protein	8	2
Kcr_0092	+	92551	93837	429	aminotransferase, class I and II			arCOG00492	ARO8	K	Transcriptional regulators containing a DNA-binding HTH domain and an aminotransferase domain (MocR family)	11	1
Kcr_0093	+	93967	94845	293	hydro-lyases, Fe-S type, tartrate/fumarate subfamily, alpha region	ttdA_fumA_fumB		arCOG04407	TtdA	C	Tartrate dehydratase alpha subunit/Fumarate hydratase class I, N-terminal domain	6	2
Kcr_0094	+	94858	95457	200	hydro-lyases, Fe-S type, tartrate/fumarate subfamily, beta region	ttdB_fumA_fumB		arCOG04406	FumA	C	Tartrate dehydratase beta subunit/Fumarate hydratase class I, C-terminal domain	6	2
Kcr_0095	-	95865	97259	465	FAD dependent oxidoreductase			arCOG02231	-	R	Uncharacterized FAD-dependent dehydrogenase	8	1
Kcr_0096	-	97352	98041	230	conserved hypothetical protein			arCOG01150	-	R	Predicted ICC-like phosphoesterase	13	2
Kcr_0097	+	98044	98418	125	hypothetical protein			arCOG01120	-	S	Uncharacterized metal-binding protein	8	0
Kcr_0098	+	98455	99234	260	helix-turn-helix motif			arCOG04060	-	K	Predicted transcriptional regulator	12	2

Gene	Strand	Start	Stop	Size (a.a.)	Product	Gene Name	COGs	arCOGs	COG_gene	Func	arCOG annotation	# of Cren in arCOGs	#
Kcr_0099	-	99507	100376	290	hypothetical protein								
Kcr_0100	+	100481	101773	431	hypothetical protein								
Kcr_0101	-	101912	102787	292	hypothetical protein								
Kcr_0102	+	102876	104072	399	hypothetical protein								
Kcr_0103	-	104162	105040	293	hypothetical protein								
Kcr_0104	+	105109	106353	415	hypothetical protein								
Kcr_0105	+	106593	108752	720	peptidase M28			arCOG02960/1	-	R	Predicted aminopeptidase, Iap family	10	0
Kcr_0106	-	108761	109588	276	protein of unknown function DUF6, transmembrane			arCOG00271	RhaT	G	permease of the drug/metabolite transporter (DMT) superfamily	11	2
Kcr_0107	+	109914	110732	273	hypothetical protein			arCOG02079/3	-	M	S-layer domain	1	1
Kcr_0108	+	110765	112111	449	TIP49-like			arCOG04081	TIP49	K	DNA helicase TIP49, TBP-interacting protein	12	6
Kcr_0109	+	112162	113652	497	hypothetical protein								
Kcr_0110	+	113657	114283	209	hypothetical protein			arCOG04359	-	R	Predicted RNA-binding protein containing a C-terminal EMAP domain	2	8
Kcr_0111	-	114278	115888	537	phosphoadenosine phosphosulfate reductase			arCOG00073	CysH	E	PAPS reductase related enzyme fused to RNA-binding PUA domain and ferredoxin domain	8	1
Kcr_0112	-	116005	116244	80	hypothetical protein			arCOG00393/3	-	K	Predicted membrane-associated trancriptional regulator	9	0
Kcr_0113	+	116395	116691	99	hypothetical protein								
Kcr_0114	+	116737	118671	645	3-hydroxyacyl-CoA dehydrogenase, NAD-binding			arCOG00249	FadB	I	3-hydroxyacyl-CoA dehydrogenase	10	7
Kcr_0115	-	118883	119575	231	acetyl-CoA synthetase I (NDP forming), beta subunit			arCOG01338	-	C	Acyl-CoA synthetase, ATP-grasp containing subunit	9	1
Kcr_0116	-	119654	120913	420	hypothetical protein			arCOG00472	0	R	Predicted ATPase	0	9
Kcr_0117	+	121030	123249	740	translation elongation factor aEF-2	aEF-2		arCOG01559	FusA	J	Translation elongation factor G, EF-G (GTPase)	13	2
Kcr_0118	-	123244	123813	190	CDP-alcohol phosphatidyltransferase			arCOG00670	PgsA	I	Phosphatidylglycerophosphate synthase	13	2
Kcr_0119	+	123869	124609	247	tRNA/rRNA methyltransferase (SpoU)			arCOG01018	LasT	J	rRNA methylase	7	2
Kcr_0120	+	124716	125111	132	Ribosomal protein L10E			arCOG04113/2	RplP	J	Ribosomal protein L16	13	2
Kcr_0121	+	125119	125700	194	putative RNA methylase			arCOG00910	-	J	Predicted RNA methylase	13	2
Kcr_0122	+	125697	126302	202	RNA-binding protein (consists of S1 domain and a Zn-ribbon domain)-like			arCOG00676	-	J	Predicted RNA-binding protein (consists of S1 domain and a Zn-ribbon domain)	13	1
Kcr_0123	+	126337	126882	182	conserved hypothetical protein			arCOG00908	-	S	Uncharacterized protein conserved in archaea	9	8
Kcr_0124	+	126872	127168	99	RNA polymerase, dimerisation			arCOG04111	RPB11	K	DNA-directed RNA polymerase, subunit L	13	2
Kcr_0125	-	127162	127959	266	conserved hypothetical protein			arCOG04215	-	R	NH2-acetyltransferase	3	3
Kcr_0126	-	127994	128629	212	alkaline phosphatase-like protein			arCOG03117/3	DedA	S	Uncharacterized membrane-associated protein, DedA family	5	8
Kcr_0127	-	128664	129344	227	regulatory protein, ArsR			arCOG01686/3	ArsR	K	Transcriptional regulator containing HTH domain, ArsR family	0	1
Kcr_0128	-	129352	129978	209	hypothetical protein								
Kcr_0129	-	129975	130628	218	hypothetical protein								
Kcr_0130	+	130790	131119	110	hypothetical protein								
Kcr_0131	+	131119	132264	382	aspartate kinase	aspK		arCOG00861	LysC	E	Aspartokinase	11	2
Kcr_0132	+	132261	133304	348	aspartate-semialdehyde dehydrogenase	asd_EA		arCOG00494	Asd	E	Aspartate-semialdehyde dehydrogenase	10	2
Kcr_0133	+	133339	133872	178	Deoxycytidine triphosphate deaminase			arCOG04048	Dcd	F	Deoxycytidine deaminase	12	2
Kcr_0134	+	134095	136725	877	Ribonucleoside-diphosphate reductase, adenosylcobalamin-dependent			arCOG03713/1	NrdA	F	Ribonucleotide reductase associated ATP-cone domain	1	2
Kcr_0135	+	136728	137228	167	cob(I)alamin adenosyltransferase	cobA		arCOG04678	BtuR	H	ATP:corrinoid adenosyltransferase	6	1
Kcr_0136	+	137454	137843	130	NADH-ubiquinone/plastoquinone oxidoreductase, chain 3			arCOG01557/3	NuoA	C	NADH dehydrogenase subunit A	12	1
Kcr_0137	+	137828	138352	175	NADH-quinone oxidoreductase, B subunit	nuoB		arCOG01554	NuoB	C	F420H2 dehydrogenase subunit, related to NADH:ubiquinone oxidoreductase 20 kD subunit	13	1
Kcr_0138	+	138342	138812	157	NADH dehydrogenase (ubiquinone), 30 kDa subunit			arCOG01551	NuoC	C	NADH dehydrogenase subunit C	12	1
Kcr_0139	+	138809	139978	390	NADH dehydrogenase (ubiquinone)			arCOG01548	NuoD	C	NADH dehydrogenase subunit D	13	1
Kcr_0140	+	139978	140979	334	respiratory-chain NADH dehydrogenase, subunit 1			arCOG01546	NuoH	C	NADH dehydrogenase subunit H	13	1
Kcr_0141	+	140969	141460	164	NADH-quinone oxidoreductase, chain I	NuoI		arCOG01543	NuoI	C	NADH dehydrogenase subunit I	13	2
Kcr_0142	+	141444	141926	161	NADH-ubiquinone/plastoquinone oxidoreductase, chain 6			arCOG04654/1	NuoJ	C	NADH dehydrogenase subunit J	7	1
Kcr_0143	+	141929	142234	102	NADH-ubiquinone oxidoreductase, chain 4L			arCOG03073	NuoK	C	NADH dehydrogenase subunit 4L (K,kappa)	6	1
Kcr_0144	+	142239	143762	508	proton-translocating NADH-quinone oxidoreductase, chain M	NDH_I_M		arCOG01537	NuoM	C	NADH:ubiquinone oxidoreductase subunit 4 (chain M)	11	2
Kcr_0145	+	143768	145831	688	proton-translocating NADH-quinone oxidoreductase, chain L	NDH_I_L		arCOG01539	NuoL	C	NADH dehydrogenase subunit L	12	1
Kcr_0146	+	145831	147222	464	NADH dehydrogenase (quinone)			arCOG01540	NuoN	C	NADH dehydrogenase subunit N	12	1
Kcr_0147	+	147346	147903	186	hypothetical protein								
Kcr_0148	+	147976	148587	204	protein of unknown function UPF0153			arCOG02579/1	-	R	Predicted Fe-S-cluster oxidoreductase	9	2
Kcr_0149	-	148714	149343	210	hypothetical protein			arCOG05906	0	S	Uncharacterized conserved protein	4	0
Kcr_0150	+	149352	150827	492	Pyruvate kinase			arCOG04120	PykF	G	Pyruvate kinase	12	2
Kcr_0151	-	150779	151999	407	dihydroorotase, multifunctional complex type	pyrC		arCOG00689	PyrC	F	Dihydroorotase or related cyclic amidohydrolase	13	2
Kcr_0152	-	151980	152771	264	oxidoreductase FAD/NAD(P)-binding			arCOG02199	UbiB	H	2-polyprenylphenol hydroxylase or related flavodoxin oxidoreductase	7	2
Kcr_0153	-	152776	153249	158	aspartate carbamoyltransferase, regulatory subunit	atcR		arCOG04229	PyrI	F	Aspartate carbamoyltransferase, regulatory subunit	13	2
Kcr_0154	-	153246	154136	297	aspartate carbamoyltransferase	atc		arCOG00911	PyrB	F	Aspartate carbamoyltransferase, catalytic chain	13	2
Kcr_0155	-	154133	154717	195	orotate phosphoribosyltransferase	pyrE		arCOG00029	PyrE	F	Orotate phosphoribosyltransferase	13	2
Kcr_0156	-	154704	155252	183	Orotidine 5'-phosphate decarboxylase			arCOG00081	PyrF	F	Orotidine-5'-phosphate decarboxylase	13	2
Kcr_0157	-	155402	156736	445	Adenylosuccinate synthase			arCOG04387	PurA	F	Adenylosuccinate synthase	9	2
Kcr_0158	-	156928	158799	624	protein of unknown function DUF224, cysteine-rich region			arCOG00332	GlpC	C	Membrane associated Fe-S oxidoreductase	6	7
Kcr_0159	+	158867	160138	424	FAD dependent oxidoreductase			arCOG00755	DadA	E	Glycine/D-amino acid oxidase (deaminating)	12	1
Kcr_0160	+	160221	160910	230	hypothetical protein								
Kcr_0161	+	161010	161585	192	hypothetical protein								
Kcr_0162	-	161596	163011	472	histone acetyltransferase, ELP3 family	ELP3		arCOG01361	ELP3	K	Histone acetyltransferase	10	2
Kcr_0163	-	163052	164419	456	argininosuccinate lyase	argH		arCOG01748	ArgH	E	Argininosuccinate lyase	10	2
Kcr_0164	-	164391	165629	413	Argininosuccinate synthase			arCOG00112	ArgG	E	Argininosuccinate synthase	10	2
Kcr_0165	-	165705	167051	449	hypothetical protein			arCOG00619/2	GltB	E	Glutamate synthase domain 2	9	1
Kcr_0166	-	167048	168166	373	tRNA synthetase, class II (D, K and N)			arCOG00409	AsnS	J	Aspartyl/asparaginyl-tRNA synthetase	7	7
Kcr_0167	-	168317	169768	484	Radical SAM			arCOG01356	-	C	Radical SAM superfamily enzyme	9	2
Kcr_0168	+	169845	170210	122	hypothetical protein								
Kcr_0169	+	170200	170583	128	hypothetical protein								
Kcr_0170	+	170601	171389	263	phage integrase			arCOG01241	XerC	L	Integrase	8	2
Kcr_0171	+	171451	171996	182	flavin reductase-like, FMN-binding			arCOG02017	-	R	Conserved protein/domain typically associated with flavoprotein oxygenase, DIM6/NTAB family	11	1
Kcr_0172	+	172077	173348	424	hypothetical protein		COG0386			O			
Kcr_0173	-	173345	174343	333	hypothetical protein								
Kcr_0174	+	174698	176251	518	hypothetical protein								
Kcr_0175	-	176228	177133	302	D-isomer specific 2-hydroxyacid dehydrogenase, NAD-binding			arCOG01754	SerA	H	Phosphoglycerate dehydrogenase or related dehydrogenase	12	2
Kcr_0176	+	177901	178434	178	hypothetical protein								
Kcr_0177	-	178378	178884	169	Dual specificity protein phosphatase			arCOG03413	CDC14	T	Predicted protein-tyrosine phosphatase	11	1
Kcr_0178	-	178890	179270	127	CopG-like DNA-binding			arCOG01008	NikR	K	Transcriptional regulator, CopG/Arc/MetJ family (DNA-binding and a metal-binding domains)	12	2
Kcr_0179	-	179270	180067	266	ABC-3			arCOG01006	ZnuB	P	ABC-type Mn2+/Zn2+ transport system, permease component	9	1
Kcr_0180	-	180054	180803	250	ABC transporter related			arCOG00201	ZnuC	P	ABC-type Mn/Zn transport system, ATPase component	9	1
Kcr_0181	+	180847	181803	319	periplasmic solute binding protein			arCOG01005	LraI	P	ABC-type metal ion transport system, periplasmic component/surface adhesin	9	1
Kcr_0182	+	181832	182080	83	hypothetical protein								
Kcr_0183	+	182077	182793	239	metallophosphoesterase			arCOG01156	-	R	Calcineurin-like phosphohydrolase	4	4
Kcr_0184	+	182839	184707	623	CDP-alcohol phosphatidyltransferase			arCOG00673/1	-	M	Predicted sugar nucleotidyltransferase	3	8
Kcr_0185	+	184827	185324	166	hypothetical protein								
Kcr_0186	-	185325	185993	223	regulatory protein, ArsR			arCOG02611/3	-	S	Predicted transcriptional regulator, containd two HTH domains	3	1
Kcr_0187	+	186092	186628	179	hypothetical protein								
Kcr_0188	+	186678	187523	282	protein of unknown function DUF125, transmembrane			arCOG01096	-	S	Uncharacterized membrane protein	12	9
Kcr_0189	-	187520	188386	289	5-carboxymethyl-2-hydroxymuconate delta-isomerase			arCOG00235	MhpD	Q	2-keto-4-pentenoate hydratase/2-oxohepta-3-ene-1,7-dioic acid hydratase (catechol pathway)	11	2
Kcr_0190	-	188477	189481	335	3-isopropylmalate dehydrogenase			arCOG01163	LeuB	C	Isocitrate/isopropylmalate dehydrogenase	9	2
Kcr_0191	-	189469	189966	166	3-isopropylmalate dehydratase, small subunit			arCOG02230	LeuD	E	3-isopropylmalate dehydratase small subunit	9	2
Kcr_0192	-	189963	191117	385	homoaconitate hydratase family protein			arCOG01698	LeuC	E	3-isopropylmalate dehydratase large subunit	9	2
Kcr_0193	-	191176	192687	504	Isopropylmalate/citramalate/homocitrate synthase			arCOG02092	LeuA	E	Isopropylmalate/homocitrate/citramalate synthase	9	2
Kcr_0194	-	192817	195045	743	AAA family ATPase, CDC48 subfamily	CDC48		arCOG01308	SpoVK	O	ATPase of the AAA+ class	13	2
Kcr_0195	+	195223	195903	227	methylenetetrahydrofolate reductase			arCOG00475	MetF	E	5,10-methylenetetrahydrofolate reductase	12	1
Kcr_0196	+	195924	196772	283	Tetrahydromethanopterin S-methyltransferase subunit H-like			arCOG04336	MtrH	H	Tetrahydromethanopterin S-methyltransferase, subunit H	1	1

Gene	Strand	Start	Stop	Size (a.a.)	Product	Gene Name	COGs	arCOGs	COG_gene	Func	arCOG annotation	# of Cren in arCOGs	#
Kcr_0197	-	196787	197968	394	beta-lactamase-like			arCOG00509	FpaA	C	Uncharacterized flavoprotein	3	1
Kcr_0198	-	197973	199367	465	CoA-binding			arCOG01340	-	C	Acyl-CoA synthetase (NDP forming)	11	2
Kcr_0199	-	199481	201802	774	glycosyl transferase, family 39			arCOG00561/1	-	O	Predicted membrane-bound dolichyl-phosphate-mannose-protein mannosyltransferase	10	4
Kcr_0200	-	201804	202838	345	Dolichyl-phosphate beta-D-mannosyltransferase			arCOG00894	WcaA	M	Glycosyltransferase	9	2
Kcr_0201	+	202898	203704	269	GTP-binding			arCOG00350	-	R	Predicted GTPase	11	1
Kcr_0202	-	203668	204081	138	protein of unknown function DUF123			arCOG00463	-	L	Uri superfamily endonuclease	8	2
Kcr_0203	-	204102	204851	250	HAD-superfamily hydrolase, subfamily IIA			arCOG04221	NagD	G	Phosphatase of the HAD superfamily	5	1
Kcr_0204	-	204835	205374	180	peptidase, zinc-dependent			arCOG00458	-	R	Predicted Zn-dependent protease	9	1
Kcr_0205	-	205379	205990	204	Rhomboid-like protein			arCOG01768	GlpG	E	Membrane associated serine protease	12	1
Kcr_0206	-	205984	207099	372	FAD dependent oxidoreductase			arCOG00570	FixC	C	Dehydrogenase (flavoprotein)	13	2
Kcr_0207	-	207329	208084	252	Helix-turn-helix, type 11			arCOG01580/3	Lrp	K	Transcriptional regulator (Lrp/AsnC family)	13	2
Kcr_0208	-	208161	208955	265	protein of unknown function Met10			arCOG00033	TYW2	J	Wybutosine (yW) biosynthesis enzyme, methyltransferase	13	2
Kcr_0209	-	208952	209707	252	TatD-related deoxyribonuclease			arCOG00891	TatD	L	Mg-dependent DNase	13	1
Kcr_0210	-	209704	210579	292	PP-loop			arCOG00042	MesJ	D	tRNA(Ile)-lysidine synthase MesJ	13	2
Kcr_0211	+	210820	212265	482	hypothetical protein								
Kcr_0212	-	212249	213133	295	Endonuclease V			arCOG00929/1	Nfi	L	Deoxyinosine 3'endonuclease (endonuclease V)	12	1
Kcr_0213	+	213186	215084	633	arginyl-tRNA synthetase	argS		arCOG00487	ArgS	J	Arginyl-tRNA synthetase	13	2
Kcr_0214	-	215052	215723	224	HhH-GPD			arCOG00459	Nth	L	Predicted EndoIII-related endonuclease	13	2
Kcr_0215	-	215720	216100	127	conserved hypothetical protein			arCOG04331	-	R	Predicted thioesterase	6	1
Kcr_0216	+	216160	216957	266	zinc/iron permease			arCOG00577	-	P	Predicted divalent heavy-metal cations transporter	5	1
Kcr_0217	-	217193	217699	169	hypothetical protein								
Kcr_0218	+	218153	218590	146	HEPN			arCOG01191	-	S	HEPN domain containing protein	11	1
Kcr_0219	+	218587	218904	106	DNA polymerase, beta-like region			arCOG01199/2	-	R	Predicted nucleotidyltransferase	7	0
Kcr_0220	-	218967	219854	296	Arsenite-transporting ATPase			arCOG02849	ArsA	P	Oxyanion-translocating ATPase	7	1
Kcr_0221	+	220023	221384	454	Mannose-1-phosphate guanylyltransferase/mannose-6-phosphate isomerase			arCOG02427	{ManC}	M	Mannose-1-phosphate guanylyltransferase	0	1
Kcr_0222	+	221440	222630	397	PhoU			arCOG01963	-	S	PhoU-like domain fused to TrkA-C domain	2	1
Kcr_0223	-	222631	223503	291	electron transfer flavoprotein, alpha subunit			arCOG00447	FixB	C	Electron transfer flavoprotein, alpha subunit	9	7
Kcr_0224	-	223500	224258	253	electron transfer flavoprotein beta-subunit			arCOG00446	FixA	C	Electron transfer flavoprotein, beta subunit	9	7
Kcr_0225	-	224331	225332	334	AIR synthase related protein-like			arCOG00636	HypE	O	Hydrogenase maturation factor	12	2
Kcr_0226	-	225336	225815	160	conserved hypothetical protein			arCOG05752	-	S	Predicted membrane protein	2	4
Kcr_0227	+	225925	226446	174	regulatory protein, ArsR			arCOG01684/2	-	K	Predicted transcriptional regulator	0	2
Kcr_0228	+	226487	226993	169	heat shock protein Hsp20			arCOG01833	IbpA	O	Molecular chaperone (HSP20 family)	12	8
Kcr_0229	+	227020	229218	733	AAA family ATPase, CDC48 subfamily	CDC48		arCOG01308	SpoVK	O	ATPase of the AAA+ class	13	2
Kcr_0230	+	229267	230280	338	glyceraldehyde-3-phosphate dehydrogenase, type I	gapdh		arCOG00493/2	GapA	G	Glyceraldehyde-3-phosphate dehydrogenase/erythrose-4-phosphate dehydrogenase	13	2
Kcr_0231	-	230282	233143	954	leucyl-tRNA synthetase	leuS		arCOG00809	LeuS	J	Leucyl-tRNA synthetase	13	2
Kcr_0232	+	233637	234077	147	hypothetical protein								
Kcr_0233	+	234251	234499	83	hypothetical protein								
Kcr_0234	-	234485	235042	186	HAD-superfamily hydrolase subfamily IA, variant 3			arCOG02292	Gph	R	Predicted phosphatase	1	1
Kcr_0235	-	235039	237171	711	DNA topoisomerase I	topA		arCOG01527	TopA	L	Topoisomerase IA	13	2
Kcr_0236	+	237296	241072	1259	protein of unknown function DUF11			arCOG02497/1	0	S	S-layer protein	0	1
Kcr_0237	-	241067	242665	533	coenzyme A transferase			arCOG01986	-	I	Acyl CoA:acetate/3-ketoacid CoA transferase	5	2
Kcr_0238	+	242748	243569	274	Methyltransferase type 11			arCOG01791	SmtA	Q	SAM-dependent methyltransferase	9	1
Kcr_0239	-	243531	245732	734	protein of unknown function DUF699, ATPase putative			arCOG01951	-	R	Predicted P-loop ATPase fused to an acetyltransferase	12	9
Kcr_0240	-	245843	246670	276	dihydroorotate dehydrogenase family protein	pyrD1		arCOG00603	PyrD	F	Dihydroorotate dehydrogenase	13	2
Kcr_0241	-	246705	247184	160	chromosome segregation and condensation protein ScpB			arCOG02613	-	K	Chromosome segregation and condensation protein B	1	1
Kcr_0242	-	247177	247716	180	hypothetical protein			arCOG02610	-	L	Rec8/ScpA/Scc1-like protein (kleisin family)	1	2
Kcr_0243	-	247706	249982	759	SMC protein-like			arCOG00371	Smc	D	Chromosome segregation ATPase	1	2
Kcr_0244	-	250040	250531	164	regulatory proteins, AsnC/Lrp			arCOG01580	Lrp	K	Transcriptional regulator (Lrp/AsnC family)	13	2
Kcr_0245	+	250848	251636	263	undecaprenyl diphosphate synthase	uppS		arCOG01532	UppS	I	Undecaprenyl pyrophosphate synthase	13	2
Kcr_0246	+	251626	252474	283	hypothetical protein								
Kcr_0247	+	252471	253073	201	hypothetical protein								
Kcr_0248	-	253054	253632	193	dehydroquinase class I			arCOG02097	AroD	E	3-dehydroquinate dehydratase	10	2
Kcr_0249	-	253655	254539	295	predicted phospho-2-dehydro-3-deoxyheptonate aldolase	AroFGH		arCOG04044	FbaB	G	DhnA-type fructose-1,6-bisphosphate aldolase or related enzyme	8	2
Kcr_0250	+	254545	255627	361	glycosyl transferase, family 2			arCOG01389	-	M	Glycosyltransferase	11	2
Kcr_0251	-	255607	256041	145	CBS			arCOG00606	-	R	CBS domain	11	2
Kcr_0252	+	256175	258070	632	peptidase S16, lon-like			arCOG01937	-	R	Archaeal serine protease	9	8
Kcr_0253	+	258239	258451	71	hypothetical protein								
Kcr_0254	+	258507	259319	271	4-vinyl reductase, 4VR			arCOG01688	-	R	Predicted hydrocarbon binding protein (contains V4R domain)	2	1
Kcr_0255	+	259358	260161	268	4-vinyl reductase, 4VR			arCOG01688	-	R	Predicted hydrocarbon binding protein (contains V4R domain)	2	1
Kcr_0256	-	260182	261387	402	Saccharopine dehydrogenase			arCOG00243	LYS9	E	Saccharopine dehydrogenase or related enzyme	8	5
Kcr_0257	-	261793	263595	601	aldehyde ferredoxin oxidoreductase			arCOG00706	-	C	Aldehyde:ferredoxin oxidoreductase	9	1
Kcr_0258	+	263737	265041	435	conserved hypothetical protein								
Kcr_0259	+	265044	266378	445	amidinotransferase			arCOG03109	-	E	N-Dimethylarginine dimethylaminohydrolase	1	2
Kcr_0260	+	266445	267191	249	KaiC			arCOG01171	RAD55	T	RecA-superfamily ATPase implicated in signal transduction	12	1
Kcr_0261	+	267184	267645	154	conserved hypothetical protein			arCOG00921	-	K	Predicted transcriptional regulator	8	6
Kcr_0262	+	267949	270018	690	hypothetical protein			arCOG06147/1	0	S	Uncharacterized conserved protein	0	2
Kcr_0263	+	270015	270644	210	hypothetical protein			arCOG07532/1	0	S	Uncharacterized conserved protein	0	2
Kcr_0264	+	270645	278285	2547	hypothetical protein			arCOG02497/3	0	S	S-layer protein	0	1
Kcr_0265	+	278292	279032	247	hypothetical protein								
Kcr_0266	+	279055	280509	485	Thermostable carboxypeptidase 1			arCOG04247	-	E	Zn-dependent carboxypeptidase	11	9
Kcr_0267	-	280504	281325	274	Radical SAM			arCOG01290	SplB	L	DNA repair photolyase	12	1
Kcr_0268	+	281360	281884	175	phospholipase D/Transphosphatidylase			arCOG02041/2	Cls	I	Phosphatidylserine/phosphatidylglycerophosphate/cardiolipin synthase or related enzyme	5	6
Kcr_0269	-	282423	283061	213	hypothetical protein			arCOG01424	0	S	Uncharacterized conserved protein	7	0
Kcr_0270	-	283201	283839	213	tropomyosin			arCOG01423	-	S	Uncharacterized conserved protein containing a coiled-coil domain	4	0
Kcr_0271	+	283983	285929	649	oxygen-sensitive ribonucleoside-triphosphate reductase			arCOG04889	NrdD	F	Oxygen-sensitive ribonucleoside-triphosphate reductase	6	1
Kcr_0272	-	285750	286814	355	transposase, IS605 OrfB			arCOG00683/2	-	L	Transposase	5	9
Kcr_0273	-	286804	287238	145	Protein of unknown function DUF1225			arCOG00683/2	-	L	Transposase	5	9
Kcr_0274	+	287515	287931	139	ribonucleoside-triphosphate reductase, anaerobic-like			arCOG00952/2	PflA	O	Pyruvate-formate lyase-activating enzyme	5	1
Kcr_0275	-	287960	289858	633	aldehyde ferredoxin oxidoreductase			arCOG00706	-	C	Aldehyde:ferredoxin oxidoreductase	9	1
Kcr_0276	-	289848	290363	172	4Fe-4S ferredoxin, iron-sulfur binding			arCOG01502	HycB	C	Fe-S-cluster-containing hydrogenase component 2	1	1
Kcr_0277	+	290641	291417	259	protein of unknown function DUF81			arCOG02050	-	R	Predicted permease	13	2
Kcr_0278	-	291504	292892	463	transglutaminase-like			arCOG02167	-	E	Transglutaminase-like enzyme, putative cysteine protease	1	2
Kcr_0279	-	292917	294542	542	DNA ligase I, ATP-dependent (dnl1)	dnl1		arCOG01347	CDC9	L	ATP-dependent DNA ligase	13	2
Kcr_0280	-	294564	295646	361	Succinyl-diaminopimelate desuccinylase			arCOG01107	ArgE	E	Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase or related deacylase	13	2
Kcr_0281	+	295685	296599	305	aldo/keto reductase			arCOG01618	ARA1	R	Aldo/keto reductase, related to diketogulonate reductase	9	0
Kcr_0282	+	296632	297213	194	Protein of unknown function DUF998			arCOG02008	-	S	Predicted membrane protein	12	1
Kcr_0283	-	297164	297589	142	CoA-binding			arCOG04227	-	R	Predicted CoA-binding protein	8	1
Kcr_0284	+	297699	298445	249	purine and other phosphorylases, family 1			arCOG01324	Udp	F	Uridine phosphorylase	12	6
Kcr_0285	+	298474	299292	273	agmatinase, putative	agmatinase		arCOG01700	SpeB	E	Arginase family enzyme	13	2
Kcr_0286	-	299303	300199	299	conserved hypothetical protein			arCOG00969	-	R	Predicted hydrolase (metallo-beta-lactamase superfamily)	12	1
Kcr_0287	+	300299	301342	348	arsenical-resistance protein	acr3		arCOG02190	ACR3	P	Arsenite efflux pump ACR3 or related permease	1	1
Kcr_0288	-	301349	302128	260	cobalt transport protein			arCOG02250	CbiQ	P	ABC-type cobalt transport system, permease component CbiQ or related transporter	8	2
Kcr_0289	-	302129	302683	185	protein of unknown function DUF1393			arCOG05752	-	S	Predicted membrane protein	2	4
Kcr_0290	+	302956	303645	230	Creatininase			arCOG04536	-	R	Uncharacterized protein, putative amidase	13	1
Kcr_0291	+	303757	304959	401	hypothetical protein			arCOG02517/3	0	S	NosD-like periplasmic protein fused to PKD repeat domain	0	2
Kcr_0292	-	304956	305591	212	Amino acid ABC transporter, permease protein, 3-TM region, His/Glu/Gln/Arg/opine			arCOG01798	HisM	E	ABC-type amino acid transport system, permease component	7	1
Kcr_0293	-	305578	306318	247	ABC transporter related			arCOG00923	GlnQ	E	ABC-type polar amino acid transport system, ATPase component	7	1
Kcr_0294	-	306315	306962	216	Amino acid ABC transporter, permease protein, 3-TM region, His/Glu/Gln/Arg/opine			arCOG01798	HisM	E	ABC-type amino acid transport system, permease component	7	1

Gene	Strand	Start	Stop	Size (a.a.)	Product	Gene Name	COGs	arCOGs	COG_gene	Func	arCOG annotation	# of Cren in arCOGs	#
Kcr_0295	-	306963	307832	290	extracellular solute-binding protein, family 3			arCOG01799	HisJ	E	ABC-type amino acid transport/signal transduction system, periplasmic component/domain	7	1
Kcr_0296	+	307848	308282	145	amino acid-binding ACT			arCOG00811	IlvH	E	Acetolactate synthase, small subunit	3	4
Kcr_0297	+	308321	309682	454	FAD dependent oxidoreductase			arCOG00754	-	R	Malate/quinone oxidoreductase or related dehydrogenase	4	5
Kcr_0298	+	309724	310608	295	hypothetical protein			arCOG01293/1	HcaD	R	NAD(FAD)-dependent dehydrogenase	1	1
Kcr_0299	+	310605	312035	477	carbohydrate kinase, FGGY			arCOG00024	GlpK	C	Glycerol kinase	6	1
Kcr_0300	+	312053	313393	447	FAD linked oxidase-like			arCOG00337	GlcD	C	FAD/FMN-containing dehydrogenase	12	1
Kcr_0301	+	313398	314153	252	hypothetical protein			arCOG07320/2	GlpC	C	Fe-S oxidoreductase	4	0
Kcr_0302	-	314140	314835	232	hypothetical protein								
Kcr_0303	+	314908	316080	391	tryptophanyl-tRNA synthetase	trpS		arCOG01887	TrpS	J	Tryptophanyl-tRNA synthetase	13	2
Kcr_0304	-	316138	316668	177	NUDIX hydrolase			arCOG01073	MutT	L	NUDIX family hydrolase	11	7
Kcr_0305	-	316703	316969	89	MoaD, archaeal			arCOG00536	MoaD	H	Molybdopterin converting factor, small subunit	9	2
Kcr_0306	-	317000	317422	141	YbaK/prolyl-tRNA synthetase associated region			arCOG04332	EbsC	S	Uncharacterized conserved protein	7	1
Kcr_0307	-	317419	318150	244	conserved hypothetical protein			arCOG04489	-	Q	Acetoacetate decarboxylase related enzyme	3	1
Kcr_0308	-	318181	319089	303	inner-membrane translocator			arCOG00261	-	R	Uncharacterized ABC-type transport system, permease component	7	8
Kcr_0309	-	319086	320078	331	inner-membrane translocator			arCOG00260	-	R	ABC-type uncharacterized transport system, permease component	7	8
Kcr_0310	-	320071	321576	502	ABC transporter related			arCOG00186	-	R	ABC-type uncharacterized transport system, ATPase component	8	8
Kcr_0311	-	321674	322894	407	basic membrane lipoprotein			arCOG00258	Med	R	Uncharacterized ABC-type transport system, periplasmic component/surface lipoprotein	7	7
Kcr_0312	+	322940	323932	331	Alcohol dehydrogenase GroES-like			arCOG01459	Tdh	E	Threonine dehydrogenase or related Zn-dependent dehydrogenase	11	1
Kcr_0313	-	323934	324695	254	Radical SAM			arCOG01290/2	SplB	L	DNA repair photolyase	12	1
Kcr_0314	+	324709	325881	391	protein of unknown function DUF651, N-terminal			arCOG04269	-	S	Uncharacterized conserved protein	13	2
Kcr_0315	-	326081	327247	389	major facilitator superfamily MFS_1			arCOG00132	ProP	G	permease of the major facilitator superfamily	10	1
Kcr_0316	-	327318	328724	469	Amidohydrolase 3			arCOG00691	-	R	Predicted metal-dependent hydrolase with the TIM-barrel fold	4	1
Kcr_0317	-	328763	330526	588	hypothetical protein			arCOG06073	PckA	C	Phosphoenolpyruvate carboxykinase (ATP)	2	0
Kcr_0318	-	330574	331653	360	aminotransferase, class I and II			arCOG01130	-	E	Aspartate/tyrosine/aromatic aminotransferase	13	2
Kcr_0319	-	331664	332512	283	protein of unknown function DUF89			arCOG04410	-	R	Predicted ATP-grasp domain fused to redox center	8	2
Kcr_0320	-	332496	332900	135	peptidase M52, hydrogen uptake protein			arCOG04429	HyaD	C	Ni,Fe-hydrogenase maturation factor	7	1
Kcr_0321	-	332930	334213	428	nickel-dependent hydrogenase, large subunit			arCOG01549	FrhA	C	Coenzyme F420-reducing hydrogenase, alpha subunit	0	1
Kcr_0322	-	334210	334986	259	NADH ubiquinone oxidoreductase, 20 kDa subunit			arCOG02472	FrhG	C	Coenzyme F420-reducing hydrogenase, gamma subunit	0	1
Kcr_0323	-	334987	335871	295	Oxidoreductase FAD-binding region			arCOG02199	UbiB	H	2-polyprenylphenol hydroxylase or related flavodoxin oxidoreductase	7	2
Kcr_0324	-	335868	336878	337	4Fe-4S ferredoxin, iron-sulfur binding			arCOG05128	NapF	C	Ferredoxin	0	7
Kcr_0325	+	337452	339077	542	amino acid permease-associated region			arCOG03462	PotE	E	Amino acid transporter	3	0
Kcr_0326	+	339152	339670	173	Peptidase C56, PfpI			arCOG00769	ThiJ	R	Putative intracellular protease/amidase	10	2
Kcr_0327	-	339811	341088	426	protein of unknown function DUF107			arCOG01910	NfeD	O	Short-chain alcohol dehydrogenase	9	1
Kcr_0328	-	341085	341789	235	band 7 protein			arCOG01915	HflC	O	Membrane protease subunit, stomatin/prohibitin homolog	12	2
Kcr_0329	+	342050	343420	457	aminotransferase class-III			arCOG00918	HemL	H	Glutamate-1-semialdehyde aminotransferase	11	2
Kcr_0330	-	343421	344194	258	hypothetical protein								
Kcr_0331	-	344172	344864	231	hypothetical protein								
Kcr_0332	+	345077	346801	575	amino acid permease-associated region			arCOG03462	PotE	E	Amino acid transporter	3	0
Kcr_0333	+	346919	347878	320	conserved hypothetical protein			arCOG01145/1	-	R	Icc family phosphoesterase	11	2
Kcr_0334	+	347881	348717	279	hypothetical protein			arCOG06048	0	S	Uncharacterized conserved protein	3	0
Kcr_0335	+	349092	349862	257	hypothetical protein								
Kcr_0336	+	349876	351588	571	ABC transporter related			arCOG00188	CbiO	P	ABC-type cobalt transport system, ATPase component	9	1
Kcr_0337	+	351585	352442	286	cobalt transport protein			arCOG02250	CbiQ	P	ABC-type cobalt transport system, permease component CbiQ or related transporter	8	2
Kcr_0338	+	352445	353521	359	aminotransferase, class I and II			arCOG01130	-	E	Aspartate/tyrosine/aromatic aminotransferase	13	2
Kcr_0339	-	353509	353964	152	NUDIX hydrolase			arCOG01075	-	F	NUDIX family hydrolase	12	2
Kcr_0340	-	354004	354513	170	flavin reductase-like, FMN-binding			arCOG02017	-	R	Conserved protein/domain typically associated with flavoprotein oxygenase, DIM6/NTAB family	11	1
Kcr_0341	-	354546	355028	161	hypothetical protein			arCOG05401	0	S	Uncharacterized conserved protein	6	3
Kcr_0342	-	355162	356061	300	dihydrodipicolinate synthetase			arCOG04172	DapA	E	Dihydrodipicolinate synthase/N-acetylneuraminate lyase	10	2
Kcr_0343	+	356439	356873	145	Protein of unknown function DUF1225			arCOG00683/2	-	L	Transposase	5	9
Kcr_0344	+	356959	357822	288	transposase, IS605 OrfB			arCOG00683/2	-	L	Transposase	5	9
Kcr_0345	+	358379	359218	280	oxidoreductase FAD/NAD(P)-binding			arCOG02199	UbiB	H	2-polyprenylphenol hydroxylase or related flavodoxin oxidoreductase	7	2
Kcr_0346	+	359215	360666	484	glutamate synthase (NADPH), homotetrameric	gltA		arCOG01292	GltD	E	NADPH-dependent glutamate synthase beta chain or related oxidoreductase	9	1
Kcr_0347	+	360705	360959	85	hypothetical protein								
Kcr_0348	+	360956	361273	106	hypothetical protein								
Kcr_0349	+	361296	361769	158	conserved hypothetical protein			arCOG04032	0	S	Uncharacterized conserved protein	4	0
Kcr_0350	+	361812	362465	218	Ribonuclease HII			arCOG04121	RnhB	L	Ribonuclease HII	13	2
Kcr_0351	-	362431	362901	157	thymidylate synthase complementing protein ThyX			arCOG01883/2	THY1	F	Alternative thymidylate synthase	11	6
Kcr_0352	+	363674	363895	74	Zinc finger/thioredoxin putative			arCOG01588/1	LysW	E	lysine biosynthesis protein LysW	10	8
Kcr_0353	-	363897	364643	249	hypothetical protein			arCOG00031	Apt	F	Adenine/guanine phosphoribosyltransferase or related PRPP-binding protein	12	7
Kcr_0354	+	364751	365710	320	Pyridoxal-5'-phosphate-dependent enzyme, beta subunit			arCOG01434/2	ThrC	E	Threonine synthase	13	2
Kcr_0355	+	365793	366758	322	protein of unknown function DUF1611			arCOG02828	-	R	NAD dependent epimerase/dehydratase family	2	1
Kcr_0356	+	366755	368194	480	conserved hypothetical protein			arCOG05346	0	S	Uncharacterized conserved protein	1	4
Kcr_0357	+	368290	368673	128	GCN5-related N-acetyltransferase			arCOG00834/3	WecD	K	Acetyltransferase (GNAT) family	1	4
Kcr_0358	+	368707	368976	90	hypothetical protein								
Kcr_0359	+	369016	370041	342	Alcohol dehydrogenase, zinc-binding			arCOG01458	Qor	C	NADPH:quinone reductase or related Zn-dependent oxidoreductase	6	1
Kcr_0360	+	370046	370369	108	t-RNA-binding region			arCOG01136	ARC1	R	EMAP domain RNA-binding protein	12	8
Kcr_0361	-	370356	371246	297	dihydrodipicolinate synthetase			arCOG04172	DapA	E	Dihydrodipicolinate synthase/N-acetylneuraminate lyase	10	2
Kcr_0362	-	371284	372204	307	phosphohydrolase (DHH superfamily)-like			arCOG00423	-	R	Predicted phosphohydrolase (DHH superfamily)	7	7
Kcr_0363	-	372206	373060	285	PfkB			arCOG00016	RbsK	G	Sugar kinase, ribokinase family	13	7
Kcr_0364	-	373057	373524	156	protein of unknown function UPF0066			arCOG00761	-	S	Uncharacterized conserved protein	11	1
Kcr_0365	+	373743	374876	378	polysaccharide biosynthesis protein			arCOG02209	RfbX	R	Polysaccharide biosynthesis protein, Mvin family	4	2
Kcr_0366	+	374971	375702	244	hypothetical protein			arCOG03768/2	0	S	Uncharacterized conserved protein	7	0
Kcr_0367	+	375757	376038	94	acylphosphatase			arCOG01674	AcyP	C	Acylphosphatase	11	2
Kcr_0368	+	376211	377152	314	peptidase M48, Ste24p			arCOG01331	HtpX	O	Zn-dependent protease with chaperone function	12	2
Kcr_0369	+	377243	377578	112	conserved hypothetical protein			arCOG02734	-	S	NifX family protein	9	2
Kcr_0370	-	377582	378670	363	hypothetical protein								
Kcr_0371	-	378912	380693	594	adenine deaminase	ade		arCOG00693	AdeC	F	Adenine deaminase	2	1
Kcr_0372	-	380913	382868	652	hypothetical protein			arCOG02632/3	HsdM	V	Type I restriction-modification system methyltransferase subunit	2	1
Kcr_0373	-	382901	383137	79	hypothetical protein								
Kcr_0374	+	383265	384254	330	Radical SAM			arCOG00938	-	R	Radical SAM superfamily enzyme	13	2
Kcr_0375	-	384232	386016	595	DNA ligase I, ATP-dependent (dnl1)	dnl1		arCOG01347	CDC9	L	ATP-dependent DNA ligase	13	2
Kcr_0376	-	386013	386348	112	heat shock protein Hsp20			arCOG01832/2	IbpA	O	Molecular chaperone (HSP20 family)	11	2
Kcr_0377	-	386361	387089	243	protein of unknown function DUF75			arCOG00347	-	R	Archaeal enzyme of ATP-grasp superfamily	13	2
Kcr_0378	-	387234	388244	337	aminoacyl-tRNA synthetase, class Ib			arCOG01887/2	TrpS	J	Tryptophanyl-tRNA synthetase	13	2
Kcr_0379	-	388237	388716	160	Magnesium-dependent phosphatase-1			arCOG04046	-	R	Predicted phosphatase	6	2
Kcr_0380	-	388706	389362	219	molybdopterin binding domain			arCOG00215	CinA	R	Predicted nucleotide-utilizing enzyme related to molybdopterin-biosynthesis enzyme MoeA	11	1
Kcr_0381	-	389457	390044	196	Protein of unknown function DUF207			arCOG04156	TYW3	J	Wybutosine (yW) biosynthesis enzyme	12	9
Kcr_0382	-	390045	391208	388	SufBD			arCOG01715	sufB	O	Cysteine desulfurase activator SufB	13	2
Kcr_0383	-	391208	391957	250	ABC transporter related			arCOG04236	sufC	O	Cysteine desulfurase activator ATPase	13	2
Kcr_0384	+	392018	393565	516	protein of unknown function UPF0031			arCOG00018	-	G	Predicted sugar kinase	13	2
Kcr_0385	-	393558	396527	990	conserved hypothetical protein			arCOG02087/3	-	S	Predicted membrane protein	1	6
Kcr_0386	-	396697	397146	150	pyridoxamine 5'-phosphate oxidase-related, FMN-binding			arCOG00520	-	R	Predicted flavin-nucleotide-binding protein	1	1
Kcr_0387	+	397436	398047	204	hypothetical protein			arCOG03119	0	S	DedA family membrane protein	13	0
Kcr_0388	-	398513	399088	192	membrane protein-like			arCOG04898	-	S	Uncharacterized conserved membrane protein	0	8
Kcr_0389	+	399324	401735	804	Plasma-membrane proton-efflux P-type ATPase			arCOG01578	MgtA	P	Cation transport ATPase	1	1
Kcr_0390	-	401910	402725	272	hypothetical protein			arCOG01426	0	S	Uncharacterized conserved protein	5	0
Kcr_0391	+	403076	403771	232	beta-lactamase-like			arCOG00503	-	R	Metal-dependent hydrolase of the beta-lactamase superfamily II	6	1
Kcr_0392	+	404019	404495	159	Archaeal PaREP8			arCOG03722	0	R	HEPN domain	9	0

Gene	Strand	Start	Stop	Size (a.a.)	Product	Gene Name	COGs	arCOGs	COG_gene	Func	arCOG annotation	# of Cren in arCOGs	#
Kcr_0393	+	404623	405063	147	T4 endonuclease V (pyrimidine dimer DNA glycosylase)					L			
Kcr_0394	+	405266	405907	214	peptidase A24A, prepilin type IV			arCOG02299	0	N	Peptidase A24A, prepilin type IV	12	0
Kcr_0395	-	406183	406914	244	phosphoribosyltransferase			arCOG00031	Apt	F	Adenine/guanine phosphoribosyltransferase or related PRPP-binding protein	12	7
Kcr_0396	-	406968	407558	197	phosphoribosyltransferase			arCOG00031/3	Apt	F	Adenine/guanine phosphoribosyltransferase or related PRPP-binding protein	12	7
Kcr_0397	-	407555	408256	234	cobalt transport protein			arCOG02250	CbiQ	P	ABC-type cobalt transport system, permease component CbiQ or related transporter	8	2
Kcr_0398	-	408246	409694	483	ABC transporter related			arCOG00188	CbiO	P	ABC-type cobalt transport system, ATPase component	9	1
Kcr_0399	-	409691	410257	189	hypothetical protein			arCOG05752	-	S	Predicted membrane protein	2	4
Kcr_0400	-	410381	411481	367	Protein of unknown function DUF1297			arCOG04346	-	F	5-formaminoimidazole-4-carboxamide-1-beta-D-ribofuranosyl 5'-monophosphate synthetase (purine biosynthesis)	10	1
Kcr_0401	-	411478	412617	380	Protein of unknown function DUF1297			arCOG04346	-	F	5-formaminoimidazole-4-carboxamide-1-beta-D-ribofuranosyl 5'-monophosphate synthetase (purine biosynthesis)	10	1
Kcr_0402	+	412691	412984	98	protein of unknown function DUF77			arCOG04373	-	S	Uncharacterized conserved protein	8	1
Kcr_0403	-	412992	413645	218	B3/4			arCOG04250	-	R	Solo B3/4 domain (OB-fold DNA/RNA-binding) of Phe-aaRS-beta	10	4
Kcr_0404	+	413825	414862	346	conserved hypothetical protein			arCOG01890	AmpS	E	Leucyl aminopeptidase (aminopeptidase T)	0	8
Kcr_0405	+	414961	416268	436	proteinase inhibitor I4, serpin			arCOG04933	-	O	Serine protease inhibitor	4	5
Kcr_0406	-	416635	417015	127	hypothetical protein								
Kcr_0407	+	417461	417685	75	paREP6			arCOG03755/3	0	S	Uncharacterized conserved protein	5	0
Kcr_0408	-	417811	419832	674	FAD dependent oxidoreductase			arCOG01521/1	-	Q	Phytoene dehydrogenase or related enzyme	6	1
Kcr_0409	-	421118	421891	258	daunorubicin resistance ABC transporter membrane protein	drrB		arCOG01467	-	V	ABC-type multidrug transport system, permease component	11	2
Kcr_0410	-	421888	422877	330	daunorubicin resistance ABC transporter ATP-binding subunit	drrA		arCOG00194	CcmA	V	ABC-type multidrug transport system, ATPase component	13	2
Kcr_0411	-	422874	423323	150	transcriptional regulator PadR-like			arCOG00002	-	K	Predicted transcriptional regulator, PadR family	13	1
Kcr_0412	-	423401	423904	168	Methyltransferase type 11			arCOG01792/3	SmtA	Q	SAM-dependent methyltransferase	1	1
Kcr_0413	-	424111	424521	137	hypothetical protein			arCOG00299	0	O	DsbA family protein	4	0
Kcr_0414	+	424894	425175	94	hypothetical protein			arCOG07097/3	0	S	Uncharacterized conserved protein	1	3
Kcr_0415	+	425184	425480	99	nucleic acid-binding protein, containing PIN domain			arCOG02219/2	VapC	V	Predicted nucleic acid-binding protein, contains PIN domain	6	1
Kcr_0416	-	425602	425832	77	conserved within P.aerophilum			arCOG03744/2	-	O	Predicted aspartyl protease	5	1
Kcr_0417	+	425998	427749	584	periplasmic copper-binding			arCOG02499/3	0	S	NosD-like periplasmic protein	1	7
Kcr_0418	+	427951	428409	153	hypothetical protein			arCOG03788/3	0	V	LEA14-like dessication related protein	2	2
Kcr_0419	-	428745	430049	435	hypothetical protein								
Kcr_0420	-	430127	431836	570	amino acid permease-associated region			arCOG03462	PotE	E	Amino acid transporter	3	0
Kcr_0421	+	432550	433146	199	Lysine exporter protein (LYSE/YGGA)			arCOG01947	RhtB	E	Putative threonine efflux protein	10	1
Kcr_0422	+	433164	433340	59	conserved hypothetical protein			arCOG02270	-	S	Uncharacterized conserved small protein	2	4
CRISPR1		433485	433636	50.66666667	3 repeats "TCGATGAAAAGAGGATTGAA"								
Kcr_0423	+	433841	434866	342	CRISPR-associated protein Cas1	cas1		arCOG01452	-	V	CRISPR-associated protein Cas1	10	2
Kcr_0424	+	434871	435803	311	Protein of unknown function DUF911, endonuclease-like			arCOG04195	-	V	CRISPR-associated protein, RecB family exonuclease	8	1
Kcr_0425	+	435785	436066	94	Protein of unknown function DUF196			arCOG04194	-	V	CRISPR-associated protein Cas2	9	1
Kcr_0426	+	436047	436592	182	CRISPR-associated protein Cas4	cas4		arCOG00786	-	V	CRISPR-associated protein Cas4 (RecB family exonuclease)	10	9
Kcr_0427	-	436593	437228	212	CRISPR locus-related DNA-binding			arCOG01446	ArsR	K	Transcriptional regulator containing HTH domain, ArsR family	9	1
Kcr_0428	+	437339	437665	109	CRISPR-associated protein, PAB1685			arCOG06143	0	V	CRISPR-associated protein	1	1
Kcr_0429	+	437678	438649	324	CRISPR-associated negative autoregulator			arCOG04192	-	V	CRISPR system related protein	10	6
Kcr_0430	+	438646	439482	279	CRISPR-associated protein, MJ0382			arCOG02670	-	V	CRISPR system related protein, RAMP superfamily	4	4
Kcr_0431	+	439479	440660	394	A. fulgidus predicted coding region AF1873			arCOG06144	0	V	CRISPR-associated protein	1	1
Kcr_0432	+	440749	441207	153	HEPN			arCOG01191/3	-	S	HEPN domain containing protein	11	1
Kcr_0433	+	441204	442898	565	CRISPR-associated helicase Cas3, core			arCOG01444	-	V	CRISPR-associated helicase Cas3	11	7
Kcr_0434	+	442873	443634	254	CRISPR-associated HD			arCOG01442	-	L	Predicted HD superfamily hydrolase, possibly a nuclease	5	2
CRISPR1		444085	451681	2532.333333	120 repeats "TCGATGAAAAGAGGATTGAA"								
Kcr_0435	-	446719	447450	244	hypothetical protein								
Kcr_0436	+	448167	448427	87	hypothetical protein								
Kcr_0437	-	451702	452283	194	hypothetical protein								
Kcr_0438	-	452499	452813	105	hypothetical protein								
Kcr_0439	-	452794	453921	376	hypothetical protein			arCOG05422/3	0	R	Predicted ATPase	3	0
Kcr_0440	-	453930	454682	251	CRISPR-associated protein, TM1791			arCOG02661	-	V	CRISPR system related protein, RAMP superfamily	6	4
Kcr_0441	-	454679	455878	400	CRISPR-associated protein TM1795 family-like			arCOG03891	-	V	CRISPR system related protein, RAMP superfamily	2	3
Kcr_0442	-	455875	456810	312	hypothetical protein			arCOG02665	-	V	CRISPR system related protein, RAMP superfamily	6	0
Kcr_0443	-	456811	459612	934	CRISPR-associated protein Crm2			arCOG02666	-	V	Predicted CRISPR-associated polymerase	6	4
Kcr_0444	-	459615	460088	158	CRISPR-associated protein, TM1791.1			arCOG02656	-	V	CRISPR system related protein	3	0
Kcr_0445	-	460088	461005	306	CRISPR-associated RAMP Cmr4			arCOG02657	-	V	CRISPR system related protein, RAMP superfamily	6	7
Kcr_0446	-	461284	462021	246	conserved hypothetical protein			arCOG03847/1	0	S	Uncharacterized conserved protein	3	3
Kcr_0447	+	462188	463588	467	CRISPR-associated protein DxTHG			arCOG03433	-	V	CRISPR system related protein	9	5
Kcr_0448	-	463831	465069	413	ATPase			arCOG03167	-	R	Predicted ATPase (AAA+ superfamily)	8	1
Kcr_0449	-	465330	466274	315	hypothetical protein								
Kcr_0450	-	466349	467176	276	hypothetical protein			arCOG01426	0	S	Uncharacterized conserved protein	5	0
Kcr_0451	+	467490	468749	420	ATPase			arCOG03167	-	R	Predicted ATPase (AAA+ superfamily)	8	1
Kcr_0452	+	469208	470521	438	ATPase			arCOG03167	-	R	Predicted ATPase (AAA+ superfamily)	8	1
Kcr_0453	+	470602	470892	97	DNA polymerase, beta-like region			arCOG01205	-	R	Predicted nucleotidyltransferase	11	6
Kcr_0454	-	470918	471358	147	PilT protein, N-terminal			arCOG00729/3	-	V	PIN domain containing protein	8	0
Kcr_0455	-	471355	471579	75	hypothetical protein			arCOG07235/3	0	S	Uncharacterized conserved protein	2	0
Kcr_0456	-	472727	473401	225	conserved hypothetical protein			arCOG04343	-	S	Uncharacterized conserved protein	0	1
Kcr_0457	-	473543	473938	132	hypothetical protein								
Kcr_0458	-	474204	474725	174	paREP15, putative coiled-coil protein			arCOG03724	0	S	Uncharacterized conserved protein	8	1
Kcr_0459	-	474914	475819	302	hypothetical protein								
Kcr_0460	-	475842	476240	133	conserved hypothetical protein			arCOG04029	0	S	Uncharacterized conserved protein	1	1
Kcr_0461	-	476243	476563	107	A. fulgidus predicted coding region AF0613			arCOG03886	0	K	CopG family DNA-binding protein	1	7
Kcr_0462	-	477169	477483	105	regulatory protein, ArsR			arCOG00732	MarR	K	Transcriptional regulator, MarR family	5	9
Kcr_0463	-	477521	478111	197	hypothetical protein								
Kcr_0464	-	478187	478543	119	hypothetical protein								
Kcr_0465	-	478827	479216	130	Dinitrogenase iron-molybdenum cofactor biosynthesis			arCOG02734	-	S	NifX family protein	9	2
Kcr_0466	-	479274	479492	73	protein of unknown function DUF167			arCOG04058/2	-	S	Uncharacterized conserved protein	3	1
Kcr_0467	-	479578	480357	260	peptidase M48, Ste24p			arCOG01334/3	HtpX	O	Zn-dependent protease with chaperone function	0	5
Kcr_0468	-	480676	481107	144	HEPN			arCOG01191	-	S	HEPN domain containing protein	11	1
Kcr_0469	-	481100	481477	126	hypothetical protein			arCOG01204/3	-	R	Predicted nucleotidyltransferase	9	3
Kcr_0470	-	481576	481824	83	hypothetical protein			arCOG03755	0	S	Uncharacterized conserved protein	5	0
Kcr_0471	-	481809	482360	184	hypothetical protein			arCOG05545	0	S	Uncharacterized conserved protein	6	0
Kcr_0472	-	482361	482582	74	paREP6			arCOG03755	0	S	Uncharacterized conserved protein	5	0
Kcr_0473	-	483186	484025	280	hypothetical protein								
Kcr_0474	-	484470	484913	148	protein of unknown function DUF55			arCOG02727	-	S	Uncharacterized protein conserved in archaea	3	8
Kcr_0475	+	485508	487544	679	hypothetical protein		COG0286			L			
Kcr_0476	-	487516	488286	257	hypothetical protein								
Kcr_0477	-	488396	488542	49	conserved hypothetical protein								
Kcr_0478	+	488769	489257	163	paREP15, putative coiled-coil protein			arCOG03724	0	S	Uncharacterized conserved protein	8	1
Kcr_0479	-	489374	490147	258	1-(5-phosphoribosyl)-5-amino-4-imidazole-carboxylate (AIR) carboxylase			arCOG02465	-	R	NCAIR mutase (PurE)-related protein	1	1
Kcr_0480	+	490188	491363	392	protein of unknown function DUF111			arCOG02701	-	S	Uncharacterized conserved protein	1	1
Kcr_0481	+	491350	492132	261	conserved hypothetical protein			arCOG00043	-	R	ATP-utilizing enzyme of the PP-loop superfamily	1	1
Kcr_0482	+	492183	493076	298	ABC-type nitrate/sulfonate/bicarbonate transport systems periplasmic components-like			arCOG01803/2	TauA	P	ABC-type nitrate/sulfonate/bicarbonate transport system, periplasmic component	4	1
Kcr_0483	+	493079	493807	243	binding-protein-dependent transport systems inner membrane component			arCOG00169	TauC	P	ABC-type nitrate/sulfonate/bicarbonate transport system, permease component	4	1
Kcr_0484	+	493804	494496	231	ABC transporter related			arCOG00193	TauB	P	ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component	11	1
Kcr_0485	+	494486	495895	470	Carboxylyase-related protein			arCOG01671	UbiD	H	3-polyprenyl-4-hydroxybenzoate decarboxylase or related decarboxylase	13	2
Kcr_0486	+	495885	496697	271	4-hydroxybenzoate polyprenyltransferase, putative	ubiA		arCOG00477	UbiA	H	4-hydroxybenzoate polyprenyltransferase or related prenyltransferase	9	4
Kcr_0487	-	496653	497429	259	metallophosphoesterase			arCOG01148	ApaH	T	Serine/threonine protein phosphatase PP2A family	3	1
Kcr_0488	-	497429	498286	286	protein of unknown function DUF6, transmembrane			arCOG00271	RhaT	G	permease of the drug/metabolite transporter (DMT) superfamily	11	2

Gene	Strand	Start	Stop	Size (a.a.)	Product	Gene Name	COGs	arCOGs	COG_gene	Func	arCOG annotation	# of Cren in arCOGs	#
Kcr_0489	+	498325	499020	232	Na+-driven multidrug efflux pump-like			arCOG01731/2	NorM	V	Na+-driven multidrug efflux pump	3	2
Kcr_0490	+	499028	499612	195	multi antimicrobial extrusion protein MatE			arCOG01731/2	NorM	V	Na+-driven multidrug efflux pump	3	2
Kcr_0491	+	499652	499903	84	Cupin 2, conserved barrel			arCOG02994/3	-	S	Cupin domain containing protein	5	1
Kcr_0492	-	499900	501027	376	hypothetical protein			arCOG05561/3	0	S	Uncharacterized conserved protein	7	0
Kcr_0493	-	501027	502439	471	protein of unknown function DUF87			arCOG00282	-	L	HerA helicase	7	0
Kcr_0494	+	502487	503359	291	Abortive infection protein			arCOG02768	-	R	Metal-dependent membrane protease, CAAX family	4	9
Kcr_0495	+	503567	504022	152	regulatory proteins, AsnC/Lrp			arCOG01580	Lrp	K	Transcriptional regulator (Lrp/AsnC family)	13	2
Kcr_0496	-	504026	504307	94	hypothetical protein								
Kcr_0497	-	504308	505780	491	aldehyde dehydrogenase			arCOG01252	PutA	C	NAD-dependent aldehyde dehydrogenase	10	2
Kcr_0498	-	505777	507075	433	aminotransferase class-III			arCOG00915	GabT	E	4-aminobutyrate aminotransferase or related aminotransferase	12	1
Kcr_0499	+	507355	508128	258	conserved hypothetical protein			arCOG03639	-	R	Predicted glutamine amidotransferase	11	1
Kcr_0500	-	508125	509051	309	hypothetical protein								
Kcr_0501	+	509240	510259	340	ornithine carbamoyltransferase	otc		arCOG00912	ArgF	E	Ornithine carbamoyltransferase	13	2
Kcr_0502	+	510264	511667	468	amino acid permease-associated region			arCOG00009	PotE	E	Amino acid transporter	9	1
Kcr_0503	+	511675	512817	381	amidinotransferase			arCOG03108	ArcA	E	Arginine deiminase	4	5
Kcr_0504	-	512806	513840	345	Orn/DAP/Arg decarboxylase 2			arCOG02268	LysA	E	Diaminopimelate decarboxylase	1	2
Kcr_0505	-	513845	514777	311	Deoxyhypusine synthase			arCOG04142	DYS1	O	Deoxyhypusine synthase	13	2
Kcr_0506	-	514803	515753	317	Oligopeptide/dipeptide ABC transporter, ATP-binding protein-like			arCOG00184	AppF	E	ABC-type oligopeptide transport system, ATPase component	13	1
Kcr_0507	-	515758	516696	313	Oligopeptide/dipeptide ABC transporter, ATP-binding protein-like			arCOG00181	DppD	E	ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	13	1
Kcr_0508	-	516693	517502	270	binding-protein-dependent transport systems inner membrane component			arCOG00748	DppC	E	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	12	2
Kcr_0509	-	517513	518523	337	binding-protein-dependent transport systems inner membrane component			arCOG00751	DppB	E	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	13	2
Kcr_0510	+	518694	520247	518	extracellular solute-binding protein, family 5			arCOG01534	DdpA	E	ABC-type dipeptide transport system, periplasmic component	10	1
Kcr_0511	+	520412	521065	218	beta-lactamase-like			arCOG00504	GloB	R	Zn-dependent hydrolase, glyoxylase family	9	2
Kcr_0512	-	521083	522189	369	hypothetical protein								
CRISPR2		521470	522092	207.6666667	7 repeats "AGCTCAGCTATCTTGCTCGA"								
Kcr_0513	+	522273	524315	681	Heavy metal translocating P-type ATPase			arCOG01576	ZntA	P	Cation transport ATPase	11	2
Kcr_0514	+	524368	524928	187	Rubrerythrin			arCOG01097	-	C	Rubrerythrin	6	1
Kcr_0515	+	525245	526555	437	hypothetical protein			arCOG07491	0	S	Uncharacterized conserved protein	2	0
Kcr_0516	+	526557	527177	207	hypothetical protein			arCOG07492	0	S	Uncharacterized conserved protein	2	0
Kcr_0517	+	527271	529724	818	Formate C-acetyltransferase			arCOG06130	PflD	C	Pyruvate-formate lyase	0	3
Kcr_0518	+	529736	530635	300	Glycyl-radical enzyme activating			arCOG00952	PflA	O	Pyruvate-formate lyase-activating enzyme	5	1
Kcr_0519	+	530628	531608	327	peptidase S1 and S6, chymotrypsin/Hap			arCOG02833	DegQ	O	Serine protease Do (heat-shock protein)	10	9
Kcr_0520	-	531594	531986	131	YjgF-like protein			arCOG01630	TdcF	J	Putative translation initiation inhibitor, yjgF family	12	1
Kcr_0521	-	532018	533223	402	threonine dehydratase	ilvA		arCOG01431	IlvA	E	Threonine dehydratase	13	8
Kcr_0522	-	533274	534248	325	hypothetical protein								
Kcr_0523	-	534294	534596	101	conserved hypothetical protein								
Kcr_0524	-	534618	535004	129	Dinitrogenase iron-molybdenum cofactor biosynthesis			arCOG02734	-	S	NifX family protein	9	2
Kcr_0525	+	535073	536329	419	sodium/hydrogen exchanger			arCOG01953	KefB	P	Kef-type K+ transport system, membrane component	8	2
Kcr_0526	+	536334	536651	106	conserved hypothetical protein			arCOG04331	-	R	Predicted thioesterase	6	1
Kcr_0527	+	536932	537474	181	CBS			arCOG00600/2	-	R	CBS domain	12	2
Kcr_0528	+	537471	537734	88	hypothetical protein								
Kcr_0529	+	537724	538920	399	cysteine desulfurases, SufS subfamily	sufS		arCOG00065	csdA	O	Selenocysteine lyase/Cysteine desulfurase	8	1
Kcr_0530	-	538953	540359	469	anion transporter	dass		arCOG00238	ArsB	P	Na+/H+ antiporter NhaD or related arsenite permease	11	1
Kcr_0531	-	540803	541630	276	Abortive infection protein			arCOG02768	-	R	Metal-dependent membrane protease, CAAX family	4	9
Kcr_0532	-	541669	542751	361	Radical SAM			arCOG00947	PflA	O	Pyruvate-formate lyase-activating enzyme	6	1
Kcr_0533	-	542748	543680	311	Helix-turn-helix, type 11			arCOG00374	-	K	Predicted membrane-associated trancriptional regulator	10	4
Kcr_0534	-	543720	544751	344	hypothetical protein								
Kcr_0535	-	544795	545652	286	hypothetical protein			arCOG02079/3	-	M	S-layer domain	1	1
Kcr_0536	-	545657	546967	437	cation diffusion facilitator family transporter	CDF		arCOG01474/1	MMT1	P	Predicted Co/Zn/Cd cation transporter	9	2
Kcr_0537	-	547010	547777	256	hypothetical protein								
Kcr_0538	+	547872	548756	295	SpoVT/AbrB-like			arCOG00318	PhoU	P	Phosphate uptake regulator	13	2
Kcr_0539	+	548784	549302	173	nitroreductase			arCOG00288	NfnB	C	Nitroreductase	11	2
Kcr_0540	+	549299	549868	190	flavin reductase-like, FMN-binding			arCOG02017	-	R	Conserved protein/domain typically associated with flavoprotein oxygenase, DIM6/NTAB family	11	1
Kcr_0541	-	549824	550837	338	hypothetical protein			arCOG07790/3	0	R	D-glucuronyl C5-epimerase C-terminal domain related protein	0	2
Kcr_0542	+	550948	551940	331	daunorubicin resistance ABC transporter ATP-binding subunit	drrA		arCOG00194	CcmA	V	ABC-type multidrug transport system, ATPase component	13	2
Kcr_0543	+	551942	552700	253	ABC-2 type transporter			arCOG01467	-	V	ABC-type multidrug transport system, permease component	11	2
Kcr_0544	-	552710	553147	146	hypothetical protein								
Kcr_0545	+	553215	553679	155	hypothetical protein								
Kcr_0546	+	553782	555251	490	Radical SAM			arCOG01357	-	C	Radical SAM superfamily enzyme	12	9
Kcr_0547	-	555239	555832	198	conserved membrane protein			arCOG01573	MscS	M	Small-conductance mechanosensitive channel	9	3
Kcr_0548	-	556010	556375	122	hypothetical protein								
Kcr_0549	-	556520	557365	282	conserved hypothetical protein			arCOG04893	-	S	Uncharacterized protein conserved in archaea	0	1
Kcr_0550	-	557370	557945	192	protein of unknown function DUF340, membrane			arCOG01615/2	-	S	Predicted membrane protein	10	7
Kcr_0551	+	558265	558705	147	OsmC-like protein			arCOG03686	-	O	Predicted redox protein, regulator of disulfide bond formation	5	8
Kcr_0552	+	558733	559527	265	aldo/keto reductase			arCOG01619	ARA1	R	Aldo/keto reductase, related to diketogulonate reductase	12	1
Kcr_0553	-	559472	560119	216	molybdopterin-guanine dinucleotide biosynthesis protein A			arCOG01872	MobA	H	Molybdopterin-guanine dinucleotide biosynthesis protein A	10	2
Kcr_0554	+	560175	561533	453	Rhomboid-like protein			arCOG01768/1	GlpG	E	Membrane associated serine protease	12	1
Kcr_0555	+	561832	563052	407	conserved hypothetical protein			arCOG02047	-	S	Uncharacterized conserved protein	0	6
Kcr_0556	-	563041	563946	302	Auxin Efflux Carrier			arCOG04756	-	R	Predicted permease	2	1
Kcr_0557	-	563977	564348	124	desulfoferrodoxin ferrous iron-binding domain	dfxfe		arCOG02146	-	C	Desulfoferrodoxin	3	1
Kcr_0558	+	564592	565104	171	protein of unknown function DUF204			arCOG04898	-	S	Uncharacterized conserved membrane protein	0	8
Kcr_0559	-	565120	565509	130	hypothetical protein								
Kcr_0560	-	565530	567410	627	conserved hypothetical protein			arCOG02264/3	-	S	Predicted membrane protein	1	1
Kcr_0561	+	567523	568131	203	multiple antibiotic resistance (MarC)-related proteins			arCOG01997	MarC	U	Multiple antibiotic transporter	12	2
Kcr_0562	-	568121	569218	366	Mandelate racemase/muconate lactonizing enzyme-like			arCOG01168	-	M	L-alanine-DL-glutamate epimerase or related enzyme of enolase superfamily	9	9
Kcr_0563	-	569202	570473	424	hydroxymethylglutaryl-CoA reductase, degradative	HMG_CoA_R_NAD		arCOG04260	HMG1	I	Hydroxymethylglutaryl-CoA reductase	13	2
Kcr_0564	+	570521	571336	272	protein of unknown function DUF6, transmembrane			arCOG00272	RhaT	G	permease of the drug/metabolite transporter (DMT) superfamily	9	7
Kcr_0565	+	571429	572385	319	hypothetical protein								
Kcr_0566	+	572390	573331	314	hypothetical protein			arCOG01989/3	-	J	Predicted Zn-ribbon RNA-binding protein with a function in translation	13	2
Kcr_0567	+	573328	574797	490	glycosyl transferase, family 2			arCOG01389/1	-	M	Glycosyltransferase	11	2
Kcr_0568	+	574831	575694	288	beta-lactamase-like			arCOG00498	GloB	R	Zn-dependent hydrolase, glyoxylase family	10	1
Kcr_0569	-	575684	577423	580	hypothetical protein								
Kcr_0570	-	577408	578877	490	ATP-dependent protease La			arCOG02163	-	O	Predicted ATP-dependent Lon-type protease	3	4
Kcr_0571	-	578933	579934	334	protein of unknown function DUF354			arCOG01395	-	R	Lipid-A-disaccharide synthase related glycosyltransferase	7	1
Kcr_0572	-	579931	581136	402	protein of unknown function DUF401			arCOG04354	-	S	Uncharacterized conserved protein	5	4
Kcr_0573	-	581263	581688	142	protein of unknown function UPF0153			arCOG02579	-	R	Predicted Fe-S-cluster oxidoreductase	9	2
Kcr_0574	+	581849	584248	800	hypothetical protein								
Kcr_0575	+	584390	585598	403	acetylornithine deacetylase or succinyl-diaminopimelate desuccinylase	DapE-ArgE		arCOG01107	ArgE	E	Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase or related deacylase	13	2
Kcr_0576	+	585595	586074	160	hypothetical protein								
Kcr_0577	+	586263	587492	410	hypothetical protein								
Kcr_0578	+	587656	588648	331	ATPase			arCOG03169	-	R	Predicted ATPase (AAA+ superfamily)	9	8
Kcr_0579	-	588703	589689	329	hypothetical protein								
Kcr_0580	-	589803	590063	87	hypothetical protein			arCOG03837	LplA	H	Lipoate-protein ligase A	2	8
Kcr_0581	-	590068	590799	244	biotin/lipoate A/B protein ligase			arCOG01939	LplA	H	Lipoate-protein ligase A	6	1
Kcr_0582	-	590804	591532	243	hypothetical protein								
Kcr_0583	+	591657	592097	147	Ferritin and Dps			arCOG01094	Bfr	P	Bacterioferritin (cytochrome b1)	1	4
Kcr_0584	+	592167	592613	149	regulatory proteins, AsnC/Lrp			arCOG01580	Lrp	K	Transcriptional regulator (Lrp/AsnC family)	13	2
Kcr_0585	-	592608	593462	285	hypothetical protein								

Gene	Strand	Start	Stop	Size (a.a.)	Product	Gene Name	COGs	arCOGs	COG_gene	Func	arCOG annotation	# of Cren in arCOGs	#
Kcr_0586	+	593638	594699	354	hypothetical protein			arCOG00287	0	S	Predicted NurA-like nuclease	10	0
Kcr_0587	+	596532	597188	219	Protein splicing (intein) site			arCOG00280/2	-	L	HerA helicase	13	2
Kcr_0588	+	597210	598697	496	pyruvate carboxyltransferase			arCOG02092	LeuA	E	Isopropylmalate/homocitrate/citramalate synthase	9	2
Kcr_0589	-	598678	599781	368	Malate dehydrogenase			arCOG04874	-	C	Malate/L-lactate dehydrogenase	2	8
Kcr_0590	+	599863	600219	119	protein of unknown function DUF498			arCOG04337	-	S	Uncharacterized conserved protein	2	1
Kcr_0591	-	600204	601010	269	Sec-independent periplasmic protein translocase			arCOG01919	TatC	U	Sec-independent protein secretion pathway component TatC	12	1
Kcr_0592	-	600997	601785	263	Methyltransferase type 12			arCOG01791	SmtA	Q	SAM-dependent methyltransferase	9	1
Kcr_0593	-	601837	602070	78	hypothetical protein								
Kcr_0594	+	602139	603218	360	Uncharacterized protein linocin/CFP29 homolog-like			arCOG05908	-	S	Uncharacterized conserved protein	6	0
Kcr_0595	-	603538	603909	124	conserved hypothetical protein			arCOG00021/2	-	S	Uncharacterized conserved protein	11	1
Kcr_0596	-	603910	605184	425	PAS			arCOG01472	-	S	Hemerythrin HHE cation binding domain containing protein	1	3
Kcr_0597	+	605329	606102	258	hypothetical protein								
Kcr_0598	+	606102	607877	592	ATPase								
Kcr_0599	+	607867	609171	435	major facilitator superfamily MFS_1			arCOG00143	ProP	G	permease of the major facilitator superfamily	5	1
Kcr_0600	+	609401	610165	255	short-chain dehydrogenase/reductase SDR			arCOG01259	FabG	I	Short-chain alcohol dehydrogenase	13	1
Kcr_0601	+	610162	611535	458	Glycine dehydrogenase (decarboxylating)			arCOG00077	GcvP	E	Glycine cleavage system protein P (pyridoxal-binding), N-terminal domain	5	1
Kcr_0602	+	611540	613096	519	glycine cleavage system protein P, subunit 2			arCOG00076	GcvP	E	Glycine cleavage system protein P (pyridoxal-binding), C-terminal domain	5	1
Kcr_0603	+	613093	614220	376	ATP-NAD/AcoX kinase			arCOG01350	-	R	Predicted inorganic polyphosphate/ATP-NAD kinase	6	1
Kcr_0604	-	614363	615004	214	hypothetical protein								
Kcr_0605	-	615090	616460	457	hypothetical protein			arCOG02487/3	0	S	Cell suface protein	3	1
Kcr_0606	+	616559	617893	445	type III restriction enzyme, res subunit			arCOG00874	SSL2	K	DNA or RNA helicase of superfamily II	12	1
Kcr_0607	+	617898	619403	502	protein of unknown function DUF790			arCOG04356	-	R	Predicted nuclease of restriction endonuclease-like fold	8	1
Kcr_0608	+	619442	622129	896	TM1410 hypothetical-related protein		COG2342			G			
Kcr_0609	-	622131	622637	169	Dual specificity protein phosphatase			arCOG03413	CDC14	T	Predicted protein-tyrosine phosphatase	11	1
Kcr_0610	-	622698	622883	62	Rubredoxin-type Fe(Cys)4 protein			arCOG04391	-	C	Rubredoxin	1	1
Kcr_0611	+	623130	624278	383	transposase, IS605 OrfB			arCOG00679	-	L	Transposase	9	1
Kcr_0612	-	624584	625864	427	Citrate transporter			arCOG00238	ArsB	P	Na+/H+ antiporter NhaD or related arsenite permease	11	1
Kcr_0613	+	626175	627197	341	ATPase			arCOG03169	-	R	Predicted ATPase (AAA+ superfamily)	9	8
Kcr_0614	+	627294	628373	360	ATPase			arCOG03169	-	R	Predicted ATPase (AAA+ superfamily)	9	8
Kcr_0615	+	628864	629622	253	Radical SAM			arCOG05120	-	R	Predicted Fe-S oxidoreductase	7	1
Kcr_0616	+	629722	630102	127	hypothetical protein								
Kcr_0617	+	630574	631272	233	conserved hypothetical protein			arCOG03167/2	-	R	Predicted ATPase (AAA+ superfamily)	8	1
Kcr_0618	-	631280	631891	204	hypothetical protein								
Kcr_0619	-	631962	632225	88	hypothetical protein								
Kcr_0620	+	632513	633457	315	aldo/keto reductase			arCOG01618	ARA1	R	Aldo/keto reductase, related to diketogulonate reductase	9	0
Kcr_0621	+	633843	634412	190	hypothetical protein			arCOG03839/2	-	S	Uncharacterized conserved protein	1	7
Kcr_0622	+	634445	635041	199	hypothetical protein			arCOG04793/3	-	K	Predicted transcriptional regulator	1	5
Kcr_0623	-	635059	636330	424	archaeal ATPase			arCOG03166	-	R	Predicted ATPase (AAA+ superfamily)	7	1
Kcr_0624	-	636410	637789	460	ATPase			arCOG03166	-	R	Predicted ATPase (AAA+ superfamily)	7	1
Kcr_0625	-	637820	638629	270	conserved hypothetical protein			arCOG05738	SmtA	Q	SAM-dependent methyltransferase	2	3
Kcr_0626	+	639342	639689	116	hypothetical protein								
Kcr_0627	-	639764	640111	116	hypothetical protein			arCOG01141/3	-	R	Predicted phosphoesterase	9	2
Kcr_0628	-	640180	641097	306	NAD-dependent epimerase/dehydratase			arCOG01369	WcaG	M	Nucleoside-diphosphate-sugar epimerase	11	2
Kcr_0629	+	641173	641868	232	protein of unknown function DUF124			arCOG01907	-	S	Uncharacterized conserved protein	7	2
Kcr_0630	-	641881	642300	140	membrane-bound metal-dependent hydrolase			arCOG01744	-	R	Membrane-bound metal-dependent hydrolase	3	7
Kcr_0631	-	642364	643314	317	NAD-dependent epimerase/dehydratase			arCOG04468	WcaG	M	Nucleoside-diphosphate-sugar epimerase	1	4
Kcr_0632	-	643324	644508	395	Glycine C-acetyltransferase			arCOG00113	BioF	H	7-keto-8-aminopelargonate synthetase or related enzyme	1	1
Kcr_0633	-	644616	645167	184	Protein of unknown function DUF429			arCOG04409	-	R	Predicted nuclease (RNAse H fold)	12	4
Kcr_0634	-	645177	645731	185	hypothetical protein			arCOG05752	-	S	Predicted membrane protein	2	4
Kcr_0635	-	645709	646443	245	short-chain dehydrogenase/reductase SDR			arCOG01259	FabG	I	Short-chain alcohol dehydrogenase	13	1
Kcr_0636	-	646490	647104	205	regulatory proteins, AsnC/Lrp			arCOG01580/1	Lrp	K	Transcriptional regulator (Lrp/AsnC family)	13	2
Kcr_0637	+	647116	647520	135	conserved hypothetical protein			arCOG03333	-	S	Uncharacterized conserved protein	0	1
Kcr_0638	-	647570	648913	448	conserved hypothetical protein			arCOG02559	0	S	Cell suface protein	2	1
Kcr_0639	-	649410	650387	326	ketol-acid reductoisomerase	ilvC		arCOG04465	IlvC	E	Ketol-acid reductoisomerase	9	2
Kcr_0640	-	650393	650650	86	hypothetical protein			arCOG05680	0	S	Uncharacterized conserved protein	5	0
Kcr_0641	-	650653	652449	599	acetolactate synthase, large subunit, biosynthetic type	acolacL		arCOG01998	IlvB	E	Acetolactate synthase large subunit or other thiamine pyrophosphate-requiring enzyme	10	2
Kcr_0642	+	652488	653435	316	hypothetical protein			arCOG02407	TrxA	O	Thiol-disulfide isomerase or thioredoxin	3	0
Kcr_0643	+	653549	654931	461	MATE efflux family protein	matE		arCOG01731	NorM	V	Na+-driven multidrug efflux pump	3	2
Kcr_0644	-	654894	655655	254	hypothetical protein			arCOG04463/3	-	S	Uncharacterized conserved membrane protein	0	1
Kcr_0645	+	655802	656797	332	ABC transporter related			arCOG00194	CcmA	V	ABC-type multidrug transport system, ATPase component	13	2
Kcr_0646	+	656794	657603	270	putative ABC-2 type transport system permease protein			arCOG01465	-	V	ABC-type multidrug transport system, permease component	10	8
Kcr_0647	+	657600	658376	259	hypothetical protein								
Kcr_0648	+	658416	659471	352	Myo-inositol-1-phosphate synthase			arCOG04213	INO1	I	Myo-inositol-1-phosphate synthase	13	1
Kcr_0649	+	659543	660781	413	hypothetical protein								
Kcr_0650	-	660783	661577	265	hypothetical protein								
Kcr_0651	-	661574	662650	359	protein of unknown function DUF763			arCOG04253	-	S	Uncharacterized conserved protein	12	1
Kcr_0652	-	662690	663076	129	hypothetical protein								
Kcr_0653	-	663070	663648	193	nitroreductase			arCOG00288	NfnB	C	Nitroreductase	11	2
Kcr_0654	+	663767	664897	377	Radical SAM			arCOG00936	PflA	O	Pyruvate-formate lyase-activating enzyme	6	4
Kcr_0655	-	664872	665576	235	Haloacid dehalogenase-like hydrolase			arCOG02291	-	R	HAD superfamily hydrolase	11	2
Kcr_0656	-	665604	665897	98	conserved hypothetical protein			arCOG04849/2	0	S	Uncharacterized conserved protein	0	5
Kcr_0657	-	665915	666406	164	hypothetical protein								
Kcr_0658	-	666407	667489	361	TRAP transporter solute receptor, TAXI family			arCOG01801	Imp	R	TRAP-type uncharacterized transport system, periplasmic component	2	7
Kcr_0659	-	667572	669482	637	TRAP transporter, 4TM/12TM fusion protein			arCOG01906	-	R	TRAP-type uncharacterized transport system, fused permease component	2	7
Kcr_0661	-	670413	670817	135	protein of unknown function UPF0047			arCOG04214	-	S	Uncharacterized conserved protein	13	2
Kcr_0662	+	671015	672109	365	putative RNA methylase			arCOG00048	-	L	Predicted N6-adenine-specific DNA methylase	8	9
Kcr_0663	+	672131	672757	209	protein of unknown function DUF116			arCOG02078/2	-	S	Uncharacterized conserved protein	0	1
Kcr_0664	-	672692	674185	498	major facilitator superfamily MFS_1			arCOG00144	ProP	G	permease of the major facilitator superfamily	8	7
Kcr_0665	-	674805	675866	354	Pyridoxal-5'-phosphate-dependent enzyme, beta subunit			arCOG01434	ThrC	E	Threonine synthase	13	2
Kcr_0666	+	676000	677184	395	Transglutaminase-like			arCOG03450/2	0	R	Predicted surface protease of transglutaminase family	0	8
Kcr_0667	+	677426	679123	566	DNA methylase N-4/N-6			arCOG00115/3	-	L	DNA modification methylase	2	1
Kcr_0668	-	679179	680813	545	amino acid permease-associated region			arCOG03462	PotE	E	Amino acid transporter	3	0
Kcr_0669	-	680867	681889	341	histone deacetylase superfamily			arCOG00324	AcuC	R	Deacetylase, including yeast histone deacetylase and acetoin utilization protein	11	1
Kcr_0670	+	682078	682725	216	protein of unknown function DUF554			arCOG06126	-	R	Uncharacterized membrane protein, possible Na+ channel or pump	1	1
Kcr_0671	-	682766	684346	527	ABC transporter related			arCOG00188	CbiO	P	ABC-type cobalt transport system, ATPase component	9	1
Kcr_0672	+	684421	685317	299	glycosyl transferase, family 2			arCOG00896/2	WcaA	M	Glycosyltransferase	0	1
Kcr_0673	-	685314	685769	152	hypothetical protein								
Kcr_0674	-	686464	686748	95	hypothetical protein								
Kcr_0675	+	687226	687486	87	hypothetical protein								
Kcr_0676	+	687470	687892	141	PIN domain		COG1848			V			
Kcr_0677	-	688042	688938	299	Radical SAM			arCOG01290	SplB	L	DNA repair photolyase	12	1
Kcr_0678	-	689125	689637	171	hypothetical protein								
Kcr_0679	-	689838	690170	111	hypothetical protein								
Kcr_0680	-	690618	692768	717	periplasmic copper-binding			arCOG02497/3	0	S	S-layer protein	0	1
Kcr_0681	-	693173	693979	269	amidinotransferase			arCOG03109	-	E	N-Dimethylarginine dimethylaminohydrolase	1	2
Kcr_0682	-	694114	698262	1383	DEAD/DEAH box helicase-like			arCOG00560/3	RecQ	L	Superfamily II DNA helicase	1	8
Kcr_0683	-	698489	698887	133	nucleic acid-binding protein, containing PIN domain			arCOG02219	VapC	V	Predicted nucleic acid-binding protein, contains PIN domain	6	1
Kcr_0684	-	698884	699132	83	hypothetical protein								

Gene	Strand	Start	Stop	Size (a.a.)	Product	Gene Name	COGs	arCOGs	COG_gene	Func	arCOG annotation	# of Cren in arCOGs	#
Kcr_0686	+	701013	701303	97	Transcriptional regulator AbrB		COG2002			K			
Kcr_0687	+	701281	701520	80	hypothetical protein								
Kcr_0688	+	701520	701885	122	hypothetical protein								
Kcr_0689	+	702890	703207	106	hypothetical protein								
Kcr_0690	+	703398	703769	124	hypothetical protein			arCOG06874/3	0	S	Uncharacterized conserved protein	1	2
Kcr_0691	-	704073	704645	191	ABC-type Mn/Zn transport systems ATPase component-like			arCOG05372/2	ZnuC	P	ABC-type Mn/Zn transport system, ATPase component	6	2
Kcr_0692	+	705210	705635	142	hypothetical protein			arCOG04017	0	S	Uncharacterized conserved protein	4	0
Kcr_0693	+	705935	706183	83	hypothetical protein								
Kcr_0694	+	706197	706814	206	ABC transporter related			arCOG00194/2	CcmA	V	ABC-type multidrug transport system, ATPase component	13	2
Kcr_0695	+	706807	707547	247	hypothetical protein								
Kcr_0696	+	708048	708431	128	hypothetical protein								
Kcr_0697	+	708418	710520	701	hypothetical protein			arCOG01013/3	-	J	Protein containg a domain similar to ribosomal protein L13E	11	0
Kcr_0698	+	710501	715333	1611	DEAD/DEAH box helicase-like			arCOG06444	-	R	Distinct helicase family with a unique C-terminal domain including a metal-binding cysteine cluster	0	2
Kcr_0699	-	715493	716740	416	beta-lactamase-like			arCOG00547	-	J	mRNA degradation ribonuclease J1/J2 (metallo-beta-lactamase superfamily)	1	6
Kcr_0700	-	716737	719664	976	helicase-like			arCOG00871/3	HepA	K	Superfamily II DNA/RNA helicase, SNF2 family	4	8
Kcr_0701	-	719792	723154	1121	conserved hypothetical protein			arCOG00887	-	R	Predicted ATPase (AAA+ superfamily)	4	4
Kcr_0702	-	723159	726077	973	protein of unknown function DUF1156			arCOG00889	-	L	Adenine-specific DNA methylase containing a Zn-ribbon	4	7
Kcr_0703	-	726327	727274	316	ribonuclease HII/HIII			arCOG04116/3	-	R	ATPase (PilT family)	13	2
Kcr_0704	+	727860	728297	146	conserved hypothetical protein			arCOG04492	-	S	Uncharacterized conserved protein	0	7
Kcr_0706	-	729319	730305	329	protein of unknown function DUF114			arCOG01911	SppA	O	Periplasmic serine protease (ClpP class)	5	1
Kcr_0707	-	730307	730801	165	hypothetical protein			arCOG00724	0	K	Predicted transcriptional regulator, PadR family	2	6
Kcr_0708	+	730917	731174	86	hypothetical protein								
Kcr_0709	+	731223	731657	145	hypothetical protein								
Kcr_0710	-	732406	733896	497	Metal-dependent phosphohydrolase, HD region			arCOG04430	-	R	HD superfamily phosphohydrolase	13	2
Kcr_0711	+	734394	734963	190	GCN5-related N-acetyltransferase			arCOG00826/3	WecD	K	Acetyltransferase (GNAT) family	11	2
Kcr_0712	-	735422	736210	263	ATPase			arCOG03169/2	-	R	Predicted ATPase (AAA+ superfamily)	9	8
Kcr_0713	+	736536	736802	89	protein of unknown function DUF433			arCOG07520/1	-	S	Uncharacterized conserved protein	1	1
Kcr_0714	+	736807	737097	97	CopG-like DNA-binding		COG3609			K			
Kcr_0715	-	737352	737708	119	hypothetical protein			arCOG03744	-	O	Predicted aspartyl protease	5	1
Kcr_0716	-	737953	738363	137	PilT protein-like			arCOG00713	-	V	PIN domain containing protein	5	4
Kcr_0717	-	738347	738628	94	Transcriptional regulator AbrB			arCOG00812	AbrB	K	Transcriptional regulator AbrB	3	3
Kcr_0718	+	738866	739201	112	HEPN			arCOG01191	-	S	HEPN domain containing protein	11	1
Kcr_0719	+	739245	739640	132	DNA polymerase, beta-like region			arCOG01195/1	-	R	Predicted nucleotidyltransferase	4	7
Kcr_0720	-	739883	740245	121	paREP15, putative coiled-coil protein			arCOG03724/2	0	S	Uncharacterized conserved protein	8	1
Kcr_0721	+	740541	740816	92	hypothetical protein			arCOG06019/3	0	S	Uncharacterized conserved protein	3	0
Kcr_0722	+	740803	741153	117	conserved hypothetical protein			arCOG06020	0	S	Uncharacterized conserved protein	3	0
Kcr_0723	-	741287	741688	134	paREP15, putative coiled-coil protein			arCOG03724/2	0	S	Uncharacterized conserved protein	8	1
Kcr_0724	+	742489	742737	83	hypothetical protein								
Kcr_0725	+	742817	743806	330	histone deacetylase superfamily			arCOG00324	AcuC	R	Deacetylase, including yeast histone deacetylase and acetoin utilization protein	11	1
Kcr_0726	-	744439	745002	188	conserved hypothetical protein								
Kcr_0727	-	745436	748450	1005	PKD			arCOG02081/3	-	S	Predicted membrane protein	1	7
Kcr_0728	-	748910	749206	99	hypothetical protein								
Kcr_0729	+	749422	750165	248	Threonyl/alanyl tRNA synthetase, SAD			arCOG01254	-	R	Predicted metal-dependent hydrolase related to alanyl-tRNA synthetase HxxxH domain	11	1
Kcr_0731	+	750732	751406	225	conserved hypothetical protein			arCOG03167/2	-	R	Predicted ATPase (AAA+ superfamily)	8	1
Kcr_0732	+	751731	752603	291	Mg2+ transporter protein, CorA-like			arCOG02265	CorA	P	Mg2+ and Co2+ transporter	3	1
Kcr_0733	-	752745	753110	122	Archaeal PaREP8			arCOG03722	0	R	HEPN domain	9	0
Kcr_0734	-	753234	753533	100	protein of unknown function DUF134			arCOG02238/3	-	R	Predicted DNA-binding protein	4	1
Kcr_0735	-	753660	754418	253	pyrroline-5-carboxylate reductase	proC		arCOG00455	ProC	E	Pyrroline-5-carboxylate reductase	10	1
Kcr_0736	+	754566	754919	118	conserved hypothetical protein			arCOG05256	CDC9	L	ATP-dependent DNA ligase, N-terminal domain	0	5
Kcr_0737	+	754956	756017	354	UBA/THIF-type NAD/FAD binding fold			arCOG01676/1	ThiF	H	Dinucleotide-utilizing enzyme involved in molybdopterin and thiamine biosynthesis	13	2
Kcr_0738	-	756281	757135	285	hypothetical protein			arCOG00465	-	L	Uncharacterized protein related to Endonuclease III	7	0
Kcr_0740	-	757768	758256	163	conserved hypothetical protein			arCOG05379	-	S	Uncharacterized conserved protein	9	3
Kcr_0741	-	758279	759016	246	ABC transporter related			arCOG00193	TauB	P	ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component	11	1
Kcr_0742	-	759016	760641	542	binding-protein-dependent transport systems inner membrane component			arCOG00174	-	P	ABC-type anion transport system, duplicated permease component	9	3
Kcr_0743	-	760715	761557	281	conserved hypothetical protein								
Kcr_0744	-	761557	762339	261	KaiC			arCOG01171	RAD55	T	RecA-superfamily ATPase implicated in signal transduction	12	1
Kcr_0745	+	762461	763018	186	putative 3-methyladenine DNA glycosylase			arCOG04295	Mpg	L	3-methyladenine DNA glycosylase	3	2
Kcr_0746	+	763171	763452	94	hypothetical protein			arCOG07453	0	S	Uncharacterized conserved protein	2	0
Kcr_0747	-	763476	764180	235	Threonyl/alanyl tRNA synthetase, SAD			arCOG01254	-	R	Predicted metal-dependent hydrolase related to alanyl-tRNA synthetase HxxxH domain	11	1
Kcr_0748	+	764295	764678	128	hypothetical protein			arCOG07097/1	0	S	Uncharacterized conserved protein	1	3
Kcr_0749	+	764668	764976	103	PilT protein-like			arCOG02219	VapC	V	Predicted nucleic acid-binding protein, contains PIN domain	6	1
Kcr_0750	+	764988	765683	232	hypothetical protein			arCOG01426/2	0	S	Uncharacterized conserved protein	5	0
Kcr_0751	+	766138	766383	82	Transcriptional regulator AbrB			arCOG00824	AbrB	K	Transcriptional regulator AbrB	3	0
Kcr_0752	+	766371	766805	145	conserved hypothetical protein			arCOG02731	VapC	V	PIN domain containing protein	3	0
Kcr_0753	-	766809	767066	86	hypothetical protein								
Kcr_0754	+	767072	768037	322	glycosyl transferase, group 1			arCOG01403	RfaG	M	Glycosyltransferase	13	2
Kcr_0755	-	768611	768904	98	conserved hypothetical protein								
Kcr_0756	-	768894	769751	286	protein of unknown function DUF81			arCOG02050	-	R	Predicted permease	13	2
Kcr_0757	-	769780	770931	384	major facilitator superfamily MFS_1			arCOG00130	ProP	G	permease of the major facilitator superfamily	13	2
Kcr_0758	-	770928	772055	376	aminotransferase, class I and II			arCOG04333	-	E	Aspartate/tyrosine/aromatic aminotransferase	3	0
Kcr_0759	+	772208	772768	187	4-vinyl reductase, 4VR			arCOG01688/3	-	R	Predicted hydrocarbon binding protein (contains V4R domain)	2	1
Kcr_0760	+	772749	773867	373	regulatory protein, LysR			arCOG00230/3	-	P	Periplasmic molybdate-binding protein/domain	7	1
Kcr_0761	+	773934	774509	192	4Fe-4S ferredoxin, iron-sulfur binding			arCOG01500	HybA	C	Fe-S-cluster-containing hydrogenase component 1	11	1
Kcr_0762	+	774514	776436	641	carbon-monoxide dehydrogenase, catalytic subunit	codh		arCOG02429	-	C	6Fe-6S prismane cluster-containing protein	0	8
Kcr_0763	+	776442	777224	261	Cobyrinic acid a,c-diamide synthase			arCOG00587	CooC	D	CO dehydrogenase maturation factor	0	1
Kcr_0764	+	777230	779176	649	4Fe-4S ferredoxin, iron-sulfur binding			arCOG02235	HdrA	C	Heterodisulfide reductase, subunit A or related polyferredoxin	0	1
Kcr_0765	+	779173	780054	294	methyl-viologen-reducing hydrogenase, delta subunit			arCOG02235/3	HdrA	C	Heterodisulfide reductase, subunit A or related polyferredoxin	0	1
Kcr_0766	+	780051	780536	162	Heterodisulfide reductase subunit C-like			arCOG00964	HdrC	C	Heterodisulfide reductase, subunit C	0	1
Kcr_0767	+	780520	781269	250	protein of unknown function DUF224, cysteine-rich region			arCOG00338	HdrB	C	Heterodisulfide reductase, subunit B	3	1
Kcr_0768	+	781550	782485	312	NAD-dependent epimerase/dehydratase			arCOG01369	WcaG	M	Nucleoside-diphosphate-sugar epimerase	11	2
Kcr_0770	+	783313	783768	152	regulatory proteins, AsnC/Lrp			arCOG01580	Lrp	K	Transcriptional regulator (Lrp/AsnC family)	13	2
Kcr_0771	+	783858	784403	182	hypothetical protein								
Kcr_0772	+	784403	785452	350	Radical SAM			arCOG00938	-	R	Radical SAM superfamily enzyme	13	2
Kcr_0773	+	785452	786537	362	Radical SAM			arCOG00938	-	R	Radical SAM superfamily enzyme	13	2
Kcr_0774	+	786646	787665	340	Vitamin B6 biosynthesis protein			arCOG04075	SNZ1	H	Pyridoxine biosynthesis enzyme	12	2
Kcr_0775	+	787666	788259	198	SNO glutamine amidotransferase			arCOG00034	PDX2	H	Predicted glutamine amidotransferase involved in pyridoxine biosynthesis	12	2
Kcr_0776	+	788391	789566	392	aminotransferase, class V			arCOG00065	csdA	O	Selenocysteine lyase/Cysteine desulfurase	8	1
Kcr_0777	-	789714	790007	98	MoaD, archaeal			arCOG00536	MoaD	H	Molybdopterin converting factor, small subunit	9	2
Kcr_0778	-	790034	791134	367	Radical SAM			arCOG00938	-	R	Radical SAM superfamily enzyme	13	2
Kcr_0779	+	791232	793127	632	aldehyde ferredoxin oxidoreductase			arCOG00706	-	C	Aldehyde:ferredoxin oxidoreductase	9	1
Kcr_0780	-	793133	793594	154	4Fe-4S ferredoxin, iron-sulfur binding			arCOG01500	HybA	C	Fe-S-cluster-containing hydrogenase component 1	11	1
Kcr_0781	-	793599	794684	362	FAD-dependent pyridine nucleotide-disulphide oxidoreductase			arCOG01292/2	GltD	E	NADPH-dependent glutamate synthase beta chain or related oxidoreductase	9	1
Kcr_0782	-	794674	795237	188	Bacterio-opsin activator, HTH			arCOG02271	-	K	Predicted transcriptional regulator, c-terminal HTH-like domain	7	6
Kcr_0783	+	795299	795892	198	conserved hypothetical protein			arCOG07069	0	S	Uncharacterized conserved protein	2	0
Kcr_0784	+	795894	797264	457	hypothetical protein			arCOG07070	0	S	Uncharacterized conserved protein	2	0
Kcr_0785	-	797257	797727	157	UspA			arCOG02053	UspA	T	Nucleotide-binding protein, UspA family	12	2
Kcr_0786	+	797798	798151	118	conserved hypothetical protein			arCOG05419	0	S	Uncharacterized conserved protein	3	0
Kcr_0787	-	798172	798399	76	hypothetical protein								

Gene	Strand	Start	Stop	Size (a.a.)	Product	Gene Name	COGs	arCOGs	COG_gene	Func	arCOG annotation	# of Cren in arCOGs	#
Kcr_0788	-	798539	798790	84	hypothetical protein								
Kcr_0789	-	798787	798999	71	Protein of unknown function DUF131			arCOG02717	-	S	Predicted membrane protein	0	1
Kcr_0790	+	799274	799927	218	hypothetical protein								
Kcr_0791	+	799958	800695	246	ABC transporter related			arCOG00194/2	CcmA	V	ABC-type multidrug transport system, ATPase component	13	2
Kcr_0792	+	800704	802164	487	hypothetical protein			arCOG02444	-	R	Predicted permease	0	8
Kcr_0793	+	802491	802820	110	transcriptional regulator PadR-like			arCOG00002/3	-	K	Predicted transcriptional regulator, PadR family	13	1
Kcr_0794	-	802807	803547	247	Methyltransferase type 11			arCOG04348	UbiE	H	Methylase involved in ubiquinone/menaquinone biosynthesis	10	2
Kcr_0795	-	803549	804322	258	thiamine biosynthesis Thi4 protein			arCOG00574	0	G	Ribulose 1,5-bisphosphate synthetase, converts PRPP to RuBP, flavoprotein	12	2
Kcr_0796	+	804429	805403	325	Ornithine cyclodeaminase			arCOG01035	-	E	Alanine dehydrogenase, mu-crystallin homolog	10	1
Kcr_0797	+	805414	806847	478	Peptidase M20D, amidohydrolase			arCOG01108	AbgB	R	Metal-dependent amidase/aminoacylase/carboxypeptidase	2	1
Kcr_0798	+	807053	808366	438	Extracellular ligand-binding receptor			arCOG01021	LivK	E	ABC-type branched-chain amino acid transport system, periplasmic component	9	8
Kcr_0799	+	808487	809239	251	ABC transporter related			arCOG00925	LivG	E	ABC-type branched-chain amino acid transport system, ATPase component	9	6
Kcr_0800	+	809241	809948	236	ABC transporter related			arCOG00924	LivF	E	ABC-type branched-chain amino acid transport system, ATPase component	9	6
Kcr_0801	+	809949	810863	305	inner-membrane translocator			arCOG01269	LivH	E	Branched-chain amino acid ABC-type transport system, permease component	8	6
Kcr_0802	+	810865	811923	353	inner-membrane translocator			arCOG01273	LivM	E	ABC-type branched-chain amino acid transport system, permease component	8	5
Kcr_0803	+	811968	813011	348	proline racemase			arCOG02255/3	DapF	E	Diaminopimelate epimerase	0	1
Kcr_0804	+	813196	814836	547	Amino acid transporters-like			arCOG03462	PotE	E	Amino acid transporter	3	0
Kcr_0805	+	814893	815825	311	metallophosphoesterase			arCOG01145/1	-	R	Icc family phosphoesterase	11	2
Kcr_0806	+	815830	816636	269	conserved hypothetical protein			arCOG06048	0	S	Uncharacterized conserved protein	3	0
Kcr_0807	-	816584	817741	386	major facilitator superfamily MFS_1			arCOG00132	ProP	G	permease of the major facilitator superfamily	10	1
Kcr_0808	+	817891	818502	204	acylphosphatase			arCOG01674/1	AcyP	C	Acylphosphatase	11	2
Kcr_0809	+	818944	819960	339	N-acetyl-gamma-glutamyl-phosphate reductase	argC		arCOG00495	ArgC	E	Acetylglutamate semialdehyde dehydrogenase	10	2
Kcr_0810	+	819970	820743	258	acetylglutamate kinase	argB		arCOG00862	ArgB	E	Acetylglutamate kinase	10	2
Kcr_0811	+	820808	821002	65	conserved hypothetical protein			arCOG01588	LysW	E	lysine biosynthesis protein LysW	10	8
Kcr_0812	+	820999	821838	280	Lysine biosynthesis enzyme LysX			arCOG01589	RimK	H	Glutathione synthase/glutaminyl transferase/alpha-L-glutamate ligase	12	2
Kcr_0813	+	821835	822944	370	aminotransferase class-III			arCOG00914	ArgD	E	Ornithine/acetylornithine aminotransferase	10	2
Kcr_0814	+	822946	823962	339	N-acetyl-ornithine/N-acetyl-lysine deacetylase	dapE		arCOG01107	ArgE	E	Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase or related deacylase	13	2
Kcr_0815	+	824120	825238	373	Radical SAM			arCOG01291	SplB	L	DNA repair photolyase	8	1
Kcr_0816	-	825222	825428	69	hypothetical protein			arCOG01013/2	-	J	Protein containg a domain similar to ribosomal protein L13E	11	0
Kcr_0817	+	825557	825832	92	ATP-cone			arCOG03715	0	R	ATP-cone domain	6	1
Kcr_0818	-	825879	826454	192	isochorismatase hydrolase			arCOG01943	PncA	Q	Amidase related to nicotinamidase	12	1
Kcr_0819	-	826435	826824	130	hypothetical protein								
Kcr_0820	-	826857	827960	368	Periplasmic phosphate binding protein			arCOG00213	PstS	P	ABC-type phosphate transport system, periplasmic component	9	2
Kcr_0821	+	828126	829130	335	SpoVT/AbrB-like			arCOG00318	PhoU	P	Phosphate uptake regulator	13	2
Kcr_0822	+	829165	829971	269	short-chain dehydrogenase/reductase SDR			arCOG01259	FabG	I	Short-chain alcohol dehydrogenase	13	1
Kcr_0823	+	830148	831056	303	Integral membrane protein TerC		COG0861			P			
Kcr_0824	+	831090	832949	620	peptidase S9, prolyl oligopeptidase active site region			arCOG01646	DAP2	E	Dipeptidyl aminopeptidase/acylaminoacyl-peptidase	11	1
Kcr_0825	-	832961	833668	236	ABC transporter related			arCOG00924	LivF	E	ABC-type branched-chain amino acid transport system, ATPase component	9	6
Kcr_0826	-	833665	834387	241	ABC transporter related			arCOG00926	LivG	E	ABC-type branched-chain amino acid transport system, ATPase component	5	4
Kcr_0827	-	834374	835327	318	inner-membrane translocator			arCOG01274	LivM	E	ABC-type branched-chain amino acid transport system, permease component	5	5
Kcr_0828	-	835329	836219	297	inner-membrane translocator			arCOG01270	LivH	E	Branched-chain amino acid ABC-type transport system, permease component	5	4
Kcr_0829	-	836250	837614	455	Extracellular ligand-binding receptor			arCOG01020	LivK	E	ABC-type branched-chain amino acid transport system, periplasmic component	5	5
Kcr_0830	-	837792	839096	435	Acetylornithine transaminase			arCOG00915	GabT	E	4-aminobutyrate aminotransferase or related aminotransferase	12	1
Kcr_0831	-	839235	840695	487	serine dehydratase alpha chain		COG1760			E			
Kcr_0832	+	840781	841620	280	protein of unknown function DUF198			arCOG04301	-	S	Uncharacterized conserved protein	1	2
Kcr_0833	-	841609	842820	404	dihydropteroate synthase	dhps		arCOG01978/3	FolP	H	Dihydropteroate synthase or related enzyme	12	2
Kcr_0834	+	842874	843545	224	protein of unknown function DUF115			arCOG04303	-	R	Uncharacterized Rossmann fold enzyme	11	2
Kcr_0835	+	843536	844312	259	Uncharacterised conserved protein UCP005026			arCOG02714	-	H	GTP and metal dependent enzyme involved F420 coenzyme biosynthesis (catalyzes addition of two l-glutamates to F420 precursor)	1	2
Kcr_0836	+	844715	844933	73	paREP6			arCOG03755	0	S	Uncharacterized conserved protein	5	0
Kcr_0837	-	845200	845475	92	hypothetical protein								
Kcr_0838	-	845732	845971	80	hypothetical protein								
Kcr_0839	+	846030	846818	263	Bacitracin resistance protein BacA			arCOG04761	UppP	I	Undecaprenyl pyrophosphate phosphatase	7	9
Kcr_0840	+	846887	847123	79	hypothetical protein								
Kcr_0841	-	847181	847777	199	O-methyltransferase-like			arCOG03850	0	R	Predicted O-methyltransferase	4	1
Kcr_0842	+	848114	849322	403	glycosyl transferase, group 1			arCOG01403	RfaG	M	Glycosyltransferase	13	2
Kcr_0843	+	849303	850337	345	glycosyl transferase, family 2			arCOG01383	-	R	Predicted glycosyltransferase	5	1
Kcr_0844	+	850484	851323	280	glycosyl transferase, family 2			arCOG01383	-	R	Predicted glycosyltransferase	5	1
Kcr_0846	+	852663	853658	332	dTDP-glucose 4,6-dehydratase			arCOG01371	RfbB	M	dTDP-D-glucose 4,6-dehydratase	5	1
Kcr_0847	+	853663	854571	303	dTDP-4-dehydrorhamnose reductase			arCOG01367	RfbD	M	dTDP-4-dehydrorhamnose reductase	7	1
Kcr_0848	+	854552	855094	181	dTDP-4-dehydrorhamnose 3,5-epimerase	rmlC		arCOG04188	RfbC	M	dTDP-4-dehydrorhamnose 3,5-epimerase or related enzyme	7	1
Kcr_0849	-	855220	856737	506	polysaccharide biosynthesis protein			arCOG02209	RfbX	R	Polysaccharide biosynthesis protein, Mvin family	4	2
Kcr_0850	-	856833	857051	73	conserved hypothetical protein			arCOG04189/2	-	K	Predicted transcription regulator containing HTH domain	9	0
Kcr_0851	-	857883	858143	87	hypothetical protein			arCOG03722/2	0	R	HEPN domain	9	0
Kcr_0852	-	858309	859001	231	hypothetical protein			arCOG03721/3	0	R	HEPN domain	10	0
Kcr_0853	-	859506	860546	347	ATPase			arCOG03169	-	R	Predicted ATPase (AAA+ superfamily)	9	8
Kcr_0854	-	860850	861095	82	hypothetical protein			arCOG02731/3	VapC	V	PIN domain containing protein	3	0
Kcr_0855	-	861258	861491	78	Transcriptional regulator AbrB			arCOG00824	AbrB	K	Transcriptional regulator AbrB	3	0
Kcr_0856	+	861910	863079	390	transposase, IS605 OrfB			arCOG00679	-	L	Transposase	9	1
Kcr_0857	+	863460	863741	94	hypothetical protein			arCOG03706/2	0	S	Uncharacterized conserved protein	6	1
Kcr_0858	-	863697	863945	83	hypothetical protein								
Kcr_0859	-	864174	864683	170	hypothetical protein								
Kcr_0860	-	864780	868157	1126	hypothetical protein			arCOG02546/3	-	S	PKD repeat	0	4
Kcr_0861	-	868215	869246	344	hypothetical protein								
Kcr_0862	-	869250	869738	163	hypothetical protein								
Kcr_0863	-	869942	870589	216	hypothetical protein			arCOG02879	0	S	Uncharacterized conserved protein	2	1
Kcr_0864	-	870824	871927	368	peptidase M24			arCOG01000	PepP	E	Xaa-Pro aminopeptidase	13	2
Kcr_0865	-	871951	872286	112	Cupin 2, conserved barrel			arCOG03003/3	{ManC}	G	Mannose-6-phosphate isomerase	1	1
Kcr_0866	+	872648	873913	422	permease for cytosine/purines, uracil, thiamine, allantoin			arCOG03447	CodB	F	Purine-cytosine permease or related protein	3	8
Kcr_0867	+	873894	875246	451	Phosphomethylpyrimidine kinase type-2			arCOG00020	ThiD	H	Hydroxymethylpyrimidine/phosphomethylpyrimidine kinase	12	2
Kcr_0868	+	875243	875734	164	ThiW			arCOG04317	-	S	Predicted membrane protein	3	0
Kcr_0869	+	875715	876248	178	conserved hypothetical protein			arCOG04316	0	S	Uncharacterized conserved protein	3	0
Kcr_0870	+	876279	876995	239	membrane protein-like			arCOG02835/3	-	S	Uncharacterized conserved membrane protein	1	7
Kcr_0871	+	877041	877301	87	regulatory proteins, AsnC/Lrp			arCOG01580/2	Lrp	K	Transcriptional regulator (Lrp/AsnC family)	13	2
Kcr_0872	+	877298	877732	145	hypothetical protein			arCOG00449/2	UspA	T	Nucleotide-binding protein, UspA family	3	1
Kcr_0873	+	878152	880029	626	amino acid permease-associated region			arCOG00009/1	PotE	E	Amino acid transporter	9	1
Kcr_0874	-	880031	881092	354	ABC transporter related			arCOG00175	MalK	G	ABC-type sugar transport system, ATPase component	12	2
Kcr_0875	-	881089	881859	257	binding-protein-dependent transport systems inner membrane component			arCOG00164	CysU	O	ABC-type sulfate transport system, permease component	10	2
Kcr_0876	-	881856	882866	337	putative sulfate transport system substrate-binding protein			arCOG00219	ModA	P	ABC-type molybdate transport system, periplasmic component	8	2
Kcr_0877	+	882960	883235	92	hypothetical protein			arCOG00230/3	-	P	Periplasmic molybdate-binding protein/domain	7	1
Kcr_0878	+	883252	883986	245	putative molybdopterin biosynthesis protein MJ0886			arCOG00217/2	MoeA	H	Molybdopterin biosynthesis enzyme	10	2
Kcr_0879	-	884000	884890	297	peptidase T2, asparaginase 2			arCOG04779	-	E	Asparaginase	12	1
Kcr_0880	-	885036	886313	426	aminotransferase class-III			arCOG00915	GabT	E	4-aminobutyrate aminotransferase or related aminotransferase	12	1
Kcr_0881	+	886478	887629	384	MgtE integral membrane region			arCOG00624/1	-	P	Permease, similar to cation transporter	2	1
Kcr_0882	-	887626	889104	493	Gamma-glutamyltransferase			arCOG04053	Ggt	E	Gamma-glutamyltransferase	10	5
Kcr_0883	-	889144	890157	338	peptidase M42			arCOG01518	FrvX	G	Cellulase M or related protein	9	2
Kcr_0884	-	890416	891195	260	short-chain dehydrogenase/reductase SDR			arCOG01259	FabG	I	Short-chain alcohol dehydrogenase	13	1
Kcr_0885	-	891204	892481	426	hypothetical protein								
Kcr_0886	+	892541	893611	357	hypothetical protein								

Gene	Strand	Start	Stop	Size (a.a.)	Product	Gene Name	COGs	arCOGs	COG_gene	Func	arCOG annotation	# of Cren in arCOGs	#
Kcr_0887	+	893621	894505	295	protoheme IX farnesyltransferase	cyoE_ctaB		arCOG00479	CyoE	O	Polyprenyltransferase (cytochrome oxidase assembly factor)	12	8
Kcr_0888	+	894612	898070	1153	hypothetical protein								
Kcr_0889	+	898184	899386	401	aminotransferase, class V			arCOG00066	NifS	E	Cysteine sulfinate desulfinase/cysteine desulfurase or related enzyme	1	1
Kcr_0890	+	899435	899917	161	nitrogen-fixing NifU-like-like			arCOG02077	IscU	C	NifU homolog involved in Fe-S cluster formation	1	1
Kcr_0891	+	899922	900167	82	hypothetical protein								
Kcr_0892	+	900597	901610	338	Tubulin/FtsZ, GTPase			arCOG02201	FtsZ	D	Cell division GTPase	1	2
Kcr_0893	+	901616	902635	340	Tubulin/FtsZ, GTPase			arCOG02201	FtsZ	D	Cell division GTPase	1	2
Kcr_0894	+	902651	902953	101	hypothetical protein								
Kcr_0895	+	902965	903960	332	Tubulin/FtsZ, GTPase			arCOG02201	FtsZ	D	Cell division GTPase	1	2
Kcr_0896	+	903960	904196	79	hypothetical protein			arCOG02155	-	R	Protein implicated in RNA metabolism, contains PRC-barrel domain	0	2
Kcr_0897	+	904193	905182	330	Tubulin/FtsZ, GTPase			arCOG02201/2	FtsZ	D	Cell division GTPase	1	2
Kcr_0898	+	905186	906157	324	Tubulin/FtsZ, GTPase			arCOG02201	FtsZ	D	Cell division GTPase	1	2
Kcr_0899	+	906264	907172	303	hypothetical protein								
Kcr_0900	-	907173	907451	93	hypothetical protein								
Kcr_0901	-	907456	908886	477	hypothetical protein								
Kcr_0902	+	909227	909457	77	hypothetical protein								
Kcr_0903	-	909462	910067	202	Small GTP-binding protein domain			arCOG00362	-	R	Predicted GTPase	2	7
Kcr_0904	-	910064	911650	529	hypothetical protein								
Kcr_0905	+	911697	912452	252	hypothetical protein								
Kcr_0906	-	912481	913014	178	hypothetical protein								
Kcr_0907	+	913238	913600	121	hypothetical protein								
Kcr_0908	+	913644	914546	301	Small GTP-binding protein domain			arCOG00362/1	-	R	Predicted GTPase	2	7
Kcr_0909	+	914559	915014	152	hypothetical protein								
Kcr_0910	+	915083	915856	258	hypothetical protein								
Kcr_0911	+	915895	917559	555	dihydroxy-acid dehydratase	ilvD		arCOG04045	IlvD	E	Dihydroxyacid dehydratase/phosphogluconate dehydratase	10	2
Kcr_0912	-	917553	918011	153	hypothetical protein								
Kcr_0913	-	918040	918441	134	Cupin 2, conserved barrel			arCOG03003	{ManC}	G	Mannose-6-phosphate isomerase	1	1
Kcr_0914	-	918444	918785	114	Cupin 2, conserved barrel			arCOG02999	-	S	Cupin domain containing protein	4	2
Kcr_0915	+	918889	920382	498	hypothetical protein			arCOG04118	0	S	Uncharacterized conserved protein	3	1
Kcr_0916	+	920482	921045	188	Peptidase C56, PfpI			arCOG00769	ThiJ	R	Putative intracellular protease/amidase	10	2
Kcr_0917	+	921122	923875	918	Pyruvate, phosphate dikinase			arCOG01111	PpsA	G	Phosphoenolpyruvate synthase/pyruvate phosphate dikinase	12	2
Kcr_0918	-	923779	924540	254	ABC transporter related			arCOG00231	PstB	P	ABC-type phosphate transport system, ATPase component	9	2
Kcr_0919	-	924537	925379	281	Phosphate transport system permease protein 2			arCOG00168	PstA	P	ABC-type phosphate transport system, permease component	9	2
Kcr_0920	-	925372	926322	317	Phosphate ABC transporter, permease protein PstC			arCOG00167	PstC	P	ABC-type phosphate transport system, permease component	9	1
Kcr_0921	+	926479	928275	599	aldehyde ferredoxin oxidoreductase			arCOG00706	-	C	Aldehyde:ferredoxin oxidoreductase	9	1
Kcr_0922	-	928272	929423	384	diaminopimelate decarboxylase	lysA		arCOG02268	LysA	E	Diaminopimelate decarboxylase	1	2
Kcr_0923	-	929439	930026	196	PhoU			arCOG00232	PhoU	P	Phosphate uptake regulator	7	2
Kcr_0924	+	930280	930801	174	hypothetical protein								
Kcr_0925	+	930842	931912	357	Radical SAM			arCOG00946	PflA	O	Pyruvate-formate lyase-activating enzyme	13	2
Kcr_0926	+	931922	932995	358	glycosyl transferase, group 1			arCOG01403	RfaG	M	Glycosyltransferase	13	2
Kcr_0927	-	932971	933819	283	putative lipoprotein			arCOG04609/2	0	S	Predicted lipoprotein	1	4
Kcr_0928	-	933827	934729	301	Pyridoxal-5'-phosphate-dependent enzyme, beta subunit			arCOG01430	CysK	E	Cysteine synthase	13	1
Kcr_0929	+	935225	936718	498	extracellular solute-binding protein, family 5			arCOG01534	DdpA	E	ABC-type dipeptide transport system, periplasmic component	10	1
Kcr_0930	+	936756	937769	338	binding-protein-dependent transport systems inner membrane component			arCOG00751	DppB	E	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	13	2
Kcr_0931	+	937762	938646	295	binding-protein-dependent transport systems inner membrane component			arCOG00748	DppC	E	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	12	2
Kcr_0932	+	938650	939609	320	Oligopeptide/dipeptide ABC transporter, ATP-binding protein-like			arCOG00181	DppD	E	ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	13	1
Kcr_0933	+	939615	940586	324	Oligopeptide/dipeptide ABC transporter, ATP-binding protein-like			arCOG00184	AppF	E	ABC-type oligopeptide transport system, ATPase component	13	1
Kcr_0934	-	940707	941993	429	aminotransferase class-III			arCOG00915	GabT	E	4-aminobutyrate aminotransferase or related aminotransferase	12	1
Kcr_0935	-	942003	943127	375	Mandelate racemase/muconate lactonizing enzyme-like			arCOG01168	-	M	L-alanine-DL-glutamate epimerase or related enzyme of enolase superfamily	9	9
Kcr_0936	-	943298	945850	851	hypothetical protein			arCOG02081/3	-	S	Predicted membrane protein	1	7
Kcr_0937	-	945975	946520	182	Protein of unknown function DUF998			arCOG02008	-	S	Predicted membrane protein	12	1
Kcr_0938	-	946552	948468	639	threonyl-tRNA synthetase	thrS		arCOG00401	ThrS	J	Threonyl-tRNA synthetase	13	2
Kcr_0939	+	948755	949339	195	hypothetical protein			arCOG06019	0	S	Uncharacterized conserved protein	3	0
Kcr_0940	+	949323	949676	118	hypothetical protein			arCOG06020	0	S	Uncharacterized conserved protein	3	0
Kcr_0941	+	949957	951339	461	protein of unknown function DUF224, cysteine-rich region			arCOG00333/3	GlpC	C	Fe-S oxidoreductase	10	2
Kcr_0942	+	951350	952735	462	FAD linked oxidase-like			arCOG00337	GlcD	C	FAD/FMN-containing dehydrogenase	12	1
Kcr_0943	-	952736	954106	457	seryl-tRNA synthetase	serS		arCOG00403	SerS	J	Seryl-tRNA synthetase	13	2
Kcr_0944	-	954142	954753	204	conserved hypothetical protein			arCOG02103	0	S	Uncharacterized conserved protein	11	4
Kcr_0945	+	954796	955248	151	hypothetical protein								
Kcr_0946	-	955439	956158	240	NAD(P)H-flavin oxidoreductase			arCOG00288	NfnB	C	Nitroreductase	11	2
Kcr_0947	+	956222	957640	473	cysteinyl-tRNA synthetase	cysS		arCOG00486	CysS	J	Cysteinyl-tRNA synthetase	13	2
Kcr_0948	+	957890	959887	666	V-type H(+)-translocating pyrophosphatase	vpp		arCOG04949	OVP1	C	Inorganic pyrophosphatase	6	8
Kcr_0949	+	959936	961111	392	major facilitator superfamily MFS_1			arCOG00130	ProP	G	permease of the major facilitator superfamily	13	2
Kcr_0950	+	962014	963396	461	Malate dehydrogenase (oxaloacetate decarboxylating)			arCOG00853	SfcA	C	Malic enzyme	10	1
Kcr_0951	+	963439	965031	531	hypothetical protein			arCOG03450/3	0	R	Predicted surface protease of transglutaminase family	0	8
Kcr_0952	+	965041	966126	362	Chorismate synthase			arCOG04133	AroC	E	Chorismate synthase	10	2
Kcr_0953	+	966258	967469	404	aminotransferase, class I and II			arCOG01130	-	E	Aspartate/tyrosine/aromatic aminotransferase	13	2
Kcr_0954	+	967457	968251	265	Prephenate dehydrogenase			arCOG00245	TyrA	E	Prephenate dehydrogenase	10	2
Kcr_0955	-	968248	969213	322	Spermine synthase			arCOG00050	SpeE	E	Spermidine synthase	12	1
Kcr_0956	-	969239	970816	526	hypothetical protein								
Kcr_0957	-	970800	971597	266	Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase			arCOG00062	-	R	Predicted amidohydrolase	13	2
Kcr_0958	-	971691	971993	101	hypothetical protein								
Kcr_0959	-	972037	973434	466	hypothetical protein			arCOG02079	-	M	S-layer domain	1	1
Kcr_0960	+	973509	974231	241	Threonyl/alanyl tRNA synthetase, SAD			arCOG01254	-	R	Predicted metal-dependent hydrolase related to alanyl-tRNA synthetase HxxxH domain	11	1
Kcr_0961	+	974246	975145	300	metallophosphoesterase			arCOG01143	ApaH	T	Serine/threonine protein phosphatase PP2A family	12	7
Kcr_0962	+	975142	975795	218	Redoxin			arCOG01972/1	TrxA	O	Thiol-disulfide isomerase or thioredoxin	12	2
Kcr_0963	+	975797	976465	223	Abortive infection protein		COG1266			R			
Kcr_0964	+	976546	976989	148	protein of unknown function DUF59			arCOG01845/1	PaaD	R	Predicted metal-sulfur cluster biosynthetic enzyme	13	1
Kcr_0965	+	977079	978227	383	Saccharopine dehydrogenase			arCOG00243	LYS9	E	Saccharopine dehydrogenase or related enzyme	8	5
Kcr_0966	+	978288	978932	215	Methyltransferase type 11			arCOG04583	SmtA	Q	SAM-dependent methyltransferase	9	4
Kcr_0967	+	978969	979709	247	RecA-superfamily ATPases implicated in signal transduction-like			arCOG05482	RAD55	T	RecA-superfamily ATPase implicated in signal transduction	6	0
Kcr_0968	+	979711	980319	203	endonuclease (RecB family)-like			arCOG04306	-	L	Predicted endonuclease (RecB family)	9	1
Kcr_0969	+	980451	980912	154	YbhB and YbcL			arCOG04702	-	R	Phospholipid-binding protein	9	1
Kcr_0970	-	980909	981904	332	Isocitrate dehydrogenase (NAD+)			arCOG01163	LeuB	C	Isocitrate/isopropylmalate dehydrogenase	9	2
Kcr_0971	-	981910	982404	165	3-isopropylmalate dehydratase, small subunit			arCOG02230	LeuD	E	3-isopropylmalate dehydratase small subunit	9	2
Kcr_0972	-	982401	983642	414	3-isopropylmalate dehydratase			arCOG01698	LeuC	E	3-isopropylmalate dehydratase large subunit	9	2
Kcr_0973	-	983639	984919	427	pyruvate carboxyltransferase			arCOG02093	LeuA	E	Isopropylmalate/homocitrate/citramalate synthase	1	6
Kcr_0974	-	985087	985854	256	putative hypothetical protein			arCOG02740	-	C	Uncharacterized Fe-S protein	0	1
Kcr_0975	-	985854	987467	538	hypothetical protein			arCOG02081	-	S	Predicted membrane protein	1	7
Kcr_0976	-	987511	988797	429	Glutamate--ammonia ligase			arCOG01909	GlnA	E	Glutamine synthetase	13	2
Kcr_0977	-	988922	989503	194	phage SPO1 DNA polymerase-related protein	spo1		arCOG00905	-	L	Uracil-DNA glycosylase	13	1
Kcr_0978	-	989500	990288	263	NAD+ synthetase	nadE		arCOG00069	NadE	H	NH3-dependent NAD+ synthetase	13	2
Kcr_0979	+	990447	991034	196	hypothetical protein								
Kcr_0980	+	991031	991354	108	hypothetical protein								
Kcr_0981	-	991351	992112	254	conserved hypothetical protein			arCOG04230	-	R	HD supefamily hydrolase	7	8
Kcr_0982	-	992109	992435	109	hypothetical protein		COG0695			O			
Kcr_0983	+	992493	993917	475	RDD			arCOG01994/1	SpoIIM	S	Uncharacterized membrane protein	3	2
Kcr_0984	-	993877	994428	184	CBS			arCOG00608	-	K	Predicted transcriptional regulator with C-terminal CBS domains	2	1

Gene	Strand	Start	Stop	Size (a.a.)	Product	Gene Name	COGs	arCOGs	COG_gene	Func	arCOG annotation	# of Cren in arCOGs	#
Kcr_0985	+	994523	995569	349	Cystathionine gamma-lyase			arCOG00060	MetC	E	Cystathionine beta-lyase/cystathionine gamma-synthase	10	1
Kcr_0986	+	995665	996591	309	protein of unknown function DUF224, cysteine-rich region			arCOG00334	GlpC	C	Fe-S oxidoreductase	10	5
Kcr_0987	+	996588	997760	391	4Fe-4S ferredoxin, iron-sulfur binding			arCOG00335	-	C	Uncharacterized conserved protein containing ferredoxin-like domain	10	5
Kcr_0988	-	997804	998784	327	protein of unknown function DUF162			arCOG00335/3	-	C	Uncharacterized conserved protein containing ferredoxin-like domain	10	5
Kcr_0989	-	998831	999229	133	Mov34/MPN/PAD-1			arCOG01138	-	R	Predicted metal-dependent protease of the PAD1/JAB1 superfamily	8	8
Kcr_0990	+	999326	999670	115	hypothetical protein			arCOG03769	0	S	Uncharacterized conserved protein	7	2
Kcr_0991	-	999667	1000401	245	Translation initiation factor 2B subunit eIF-2B alpha/beta/delta family-like			arCOG01125	GCD2	J	Translation initiation factor 2B subunit, eIF-2B alpha/beta/delta family	7	5
Kcr_0992	-	1000409	1001713	435	Glycine hydroxymethyltransferase			arCOG00070	GlyA	E	Glycine/serine hydroxymethyltransferase	12	2
Kcr_0993	+	1001773	1002951	393	threonine synthase	thrC		arCOG01434	ThrC	E	Threonine synthase	13	2
Kcr_0994	+	1002948	1003841	298	homoserine kinase	thrB		arCOG01027	ThrB	E	Homoserine kinase	11	1
Kcr_0995	+	1003998	1005086	363	Radical SAM			arCOG00938	-	R	Radical SAM superfamily enzyme	13	2
Kcr_0996	-	1005487	1006155	223	protein of unknown function DUF120			arCOG01904	-	K	Transcriptional regulator of a riboflavin/FAD biosynthetic operon	13	2
Kcr_0997	-	1006145	1006783	213	Translin			arCOG04318	-	J	Predicted RNA-binding protein of the translin family	13	1
Kcr_0998	+	1006792	1007769	326	Radical SAM			arCOG04174	TYW3	C	Wybutosine (yW) biosynthesis enzyme, Fe-S oxidoreductase	12	2
Kcr_0999	+	1007779	1008459	227	Uridylate kinase, putative			arCOG00858	PyrH	F	Uridylate kinase	13	2
Kcr_1000	+	1008464	1008841	126	hypothetical protein								
Kcr_1001	+	1008949	1009500	184	thymidylate kinase related			arCOG01038	-	F	Predicted nucleotide kinase (CMP/AMP kinase related)	13	2
Kcr_1002	+	1009781	1010062	94	PRC-barrel			arCOG02155	-	R	Protein implicated in RNA metabolism, contains PRC-barrel domain	0	2
Kcr_1003	-	1010059	1010319	87	hypothetical protein								
Kcr_1004	+	1010380	1010712	111	hypothetical protein			arCOG04271	RPB8	K	DNA-directed RNA polymerase, subunit RPB8	13	0
Kcr_1005	+	1010705	1011373	223	Transcription factor TFIIB, cyclin-related			arCOG01981/2	SUA7	K	Transcription initiation factor TFIIIB, Brf1 subunit/Transcription initiation factor TFIIB	13	2
Kcr_1006	+	1011497	1013134	546	ATPase-like			arCOG00286/3	-	L	HerA helicase	5	0
Kcr_1007	+	1013580	1013921	114	hypothetical protein								
Kcr_1008	-	1014024	1015163	380	ATPase			arCOG00472/3	0	R	Predicted ATPase	0	9
Kcr_1009	+	1015267	1016265	333	hypothetical protein								
Kcr_1010	+	1016339	1017487	383	cell division control protein 6			arCOG00467	CDC6	L	Cdc6-related protein, AAA superfamily ATPase	13	2
Kcr_1011	-	1017461	1018609	383	AAA ATPase, central region			arCOG00470	HolB	L	ATPase involved in DNA replication HolB, large subunit	13	2
Kcr_1012	-	1018680	1019675	332	Replication factor C			arCOG00469	HolB	L	ATPase involved in DNA replication HolB, small subunit	13	2
Kcr_1013	+	1019739	1019942	68	RNA polymerases N/8 Kd subunits			arCOG04244	RPB10	K	DNA-directed RNA polymerase, subunit N (RpoN/RPB10)	13	2
Kcr_1014	+	1019947	1020600	218	ribosomal protein S2			arCOG04245	RpsB	J	Ribosomal protein S2	13	2
Kcr_1015	+	1020597	1021454	286	protein of unknown function DUF52			arCOG01728	-	R	Predicted dioxygenase	13	2
Kcr_1016	+	1021451	1022533	361	Isopentenyl-diphosphate delta-isomerase, FMN-dependent			arCOG00613	idi	H	Isopentenyl diphosphate isomerase	12	2
Kcr_1017	+	1022559	1022942	128	ribosomal protein L7Ae/L30e/S12e/Gadd45			arCOG01751	RPL8A	J	Ribosomal protein HS6-type (S12/L30/L7a)	13	2
Kcr_1018	+	1022989	1023204	72	ribosomal protein S28e			arCOG04314	RPS28A	J	Ribosomal protein S28E/S33	13	2
Kcr_1019	+	1023210	1023389	60	Ribosomal protein L24E			arCOG01950	RPL24A	J	Ribosomal protein L24E	13	2
Kcr_1020	+	1023400	1025190	597	translation initiation factor aIF-2	aIF-2		arCOG01560	InfB	J	Translation initiation factor 2 (IF-2; GTPase)	13	2
Kcr_1021	+	1025388	1025807	140	ribosomal protein S6e			arCOG01946	RPS6A	J	Ribosomal protein S6E (S10)	13	2
Kcr_1022	+	1025817	1027058	414	protein synthesis factor, GTP-binding			arCOG01563	GCD11	J	Translation initiation factor 2, gamma subunit (eIF-2gamma; GTPase)	13	2
Kcr_1023	-	1027055	1027729	225	Thermophilic glucose-6-phosphate isomerase and related metalloenzymes-like			arCOG02602	-	G	Thermophilic glucose-6-phosphate isomerase or related metalloenzyme	0	1
Kcr_1024	+	1027782	1029446	555	hypothetical protein								
Kcr_1025	+	1029773	1030753	327	Radical SAM			arCOG00940/2	-	R	Radical SAM superfamily enzyme	2	7
Kcr_1026	+	1030740	1031492	251	Silent information regulator protein Sir2			arCOG04248	SIR2	K	NAD-dependent protein deacetylase, SIR2 family	11	9
Kcr_1027	+	1031489	1032370	294	GCN5-related N-acetyltransferase			arCOG00844/3	WecD	K	Acetyltransferase (GNAT) family	0	1
Kcr_1028	-	1032367	1033659	431	basic membrane lipoprotein			arCOG00257	Med	R	Uncharacterized ABC-type transport system, periplasmic component/surface lipoprotein	6	2
Kcr_1029	-	1033841	1034755	305	inner-membrane translocator			arCOG00261	-	R	Uncharacterized ABC-type transport system, permease component	7	8
Kcr_1030	-	1034769	1035833	355	inner-membrane translocator			arCOG00260	-	R	ABC-type uncharacterized transport system, permease component	7	8
Kcr_1031	-	1035811	1037313	501	ABC transporter related			arCOG00186	-	R	ABC-type uncharacterized transport system, ATPase component	8	8
Kcr_1032	+	1037575	1038234	220	Peroxidase			arCOG00312	AhpC	O	Peroxiredoxin	11	2
Kcr_1033	+	1038411	1039472	354	hypothetical protein								
Kcr_1034	-	1039550	1040197	216	5-methyltetrahydrofolate--homocysteine S-methyltransferase			arCOG02028	-	R	Predicted cobalamin binding protein	2	8
Kcr_1035	-	1040211	1041326	372	Pyridoxal-5'-phosphate-dependent enzyme, beta subunit			arCOG01434	ThrC	E	Threonine synthase	13	2
Kcr_1036	-	1041348	1042565	406	acyl-CoA dehydrogenase-like			arCOG04310	CaiA	I	Acyl-CoA dehydrogenase	6	3
Kcr_1037	+	1042696	1043475	260	electron transfer flavoprotein beta-subunit			arCOG00446	FixA	C	Electron transfer flavoprotein, beta subunit	9	7
Kcr_1038	+	1043472	1044377	302	electron transfer flavoprotein, alpha subunit			arCOG00447	FixB	C	Electron transfer flavoprotein, alpha subunit	9	7
Kcr_1039	+	1044382	1045557	392	Acetyl-CoA C-acetyltransferase			arCOG01282	PaaJ	I	Acetyl-CoA acetyltransferase	9	7
Kcr_1040	+	1045849	1046175	109	molybdopterin biosynthesis MoaE			arCOG00534/2	MoaE	H	Molybdopterin converting factor, large subunit	10	1
Kcr_1041	+	1046200	1046637	146	molybdenum cofactor biosynthesis protein C	moaC		arCOG01530	MoaC	H	Molybdenum cofactor biosynthesis enzyme	12	2
Kcr_1042	+	1046624	1047556	311	Molybdenum cofactor biosynthesis protein A, archaea			arCOG00930	MoaA	H	Molybdenum cofactor biosynthesis enzyme	12	2
Kcr_1043	-	1047553	1048062	170	conserved hypothetical protein			arCOG04311	-	R	Predicted hydrolase of HD superfamily	12	9
Kcr_1044	+	1048277	1049179	301	conserved hypothetical protein			arCOG01159	-	S	Uncharacterized archaeal coiled-coil protein	13	2
Kcr_1045	+	1049212	1050294	361	Protein of unknown function DUF373			arCOG04151	-	R	ABC-type multidrug transport system, permease component	13	2
Kcr_1046	-	1050220	1050723	168	protein of unknown function DUF88			arCOG02408	-	S	Uncharacterized conserved protein	0	2
Kcr_1047	+	1050787	1051113	109	transcription factor S	TFS		arCOG00579	RPB9	K	DNA-directed RNA polymerase, subunit M/Transcription elongation factor TFIIS	13	2
Kcr_1048	-	1051110	1052381	424	3-phosphoshikimate 1-carboxyvinyltransferase	aroA		arCOG04134	AroA	E	5-enolpyruvylshikimate-3-phosphate synthase	11	2
Kcr_1049	-	1052509	1053450	314	Acetamidase/Formamidase			arCOG01004	-	C	Predicted acetamidase/formamidase	4	7
Kcr_1050	-	1053447	1055057	537	amino acid permease-associated region			arCOG03462	PotE	E	Amino acid transporter	3	0
Kcr_1051	+	1055232	1055513	94	amino acid-binding ACT			arCOG04941	-	T	ACT domain-containing protein	4	1
Kcr_1052	+	1055516	1056883	456	protein of unknown function DUF711			arCOG04321	-	S	Uncharacterized conserved protein	12	1
Kcr_1053	+	1056885	1057199	105	protein of unknown function DUF437			arCOG03398/2	-	S	Uncharacterized conserved protein	8	7
Kcr_1054	+	1057200	1058189	330	thioredoxin reductase	trxr		arCOG01296	TrxB	O	Thioredoxin reductase	13	2
Kcr_1055	-	1058152	1059210	353	protein of unknown function DUF711			arCOG04321/2	-	S	Uncharacterized conserved protein	12	1
Kcr_1056	+	1059273	1061297	675	Oligosaccharyl transferase, STT3 subunit			arCOG02044/2	-	R	Uncharacterized membrane protein, required for N-linked glycosylation	11	9
Kcr_1057	+	1061420	1062394	325	nucleoside diphosphate kinase			arCOG04313/1	Ndk	F	Nucleoside diphosphate kinase	12	2
Kcr_1058	+	1062419	1062844	142	transcriptional regulator-like			arCOG00017/1	-	R	Predicted transcriptional regulator	12	1
Kcr_1059	+	1063031	1065622	864	oligopeptide binding protein, putative			arCOG01672	-	R	Predicted solute binding protein	11	4
Kcr_1060	+	1065694	1066764	357	binding-protein-dependent transport systems inner membrane component			arCOG00751	DppB	E	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	13	2
Kcr_1061	+	1066761	1068149	463	binding-protein-dependent transport systems inner membrane component			arCOG00749	DppC	E	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	12	4
Kcr_1062	+	1068152	1069099	316	Oligopeptide/dipeptide ABC transporter, ATP-binding protein-like			arCOG00181	DppD	E	ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	13	1
Kcr_1063	+	1069105	1070064	320	Oligopeptide/dipeptide ABC transporter, ATP-binding protein-like			arCOG00184	AppF	E	ABC-type oligopeptide transport system, ATPase component	13	1
Kcr_1064	+	1070061	1070561	167	hypothetical protein								
Kcr_1065	+	1070695	1070913	73	hypothetical protein								
Kcr_1066	-	1070915	1072000	362	Myo-inositol-1-phosphate synthase			arCOG04213	INO1	I	Myo-inositol-1-phosphate synthase	13	1
Kcr_1067	+	1072340	1073278	313	mevalonate kinase	mevK		arCOG01028	ERG12	I	Mevalonate kinase	13	2
Kcr_1068	-	1073275	1074459	395	Quinolinate phosphoribosyl transferase			arCOG01481	PncB	H	Nicotinic acid phosphoribosyltransferase	12	1
Kcr_1069	+	1074663	1076093	477	hydrolase of the metallo-beta-lactamase superfamily-like			arCOG00547	-	J	mRNA degradation ribonuclease J1/J2 (metallo-beta-lactamase superfamily)	1	6
Kcr_1070	+	1076106	1077230	375	peptidase M24			arCOG01000	PepP	E	Xaa-Pro aminopeptidase	13	2
Kcr_1071	+	1077233	1078396	388	protein of unknown function Met10			arCOG00032	-	R	SAM-dependent methyltransferase	2	8
Kcr_1072	-	1078422	1078787	122	hypothetical protein								
Kcr_1073	+	1078880	1082137	1086	reverse gyrase	rgy		arCOG01526	-	L	Reverse gyrase	12	7
Kcr_1074	-	1082127	1084979	951	hypothetical protein			arCOG02556/3	-	S	WD40-like repeat	2	1
Kcr_1075	-	1085072	1086781	570	glutamyl-tRNA synthetase	gltXa		arCOG04302	GlnS	J	Glutamyl- or glutaminyl-tRNA synthetase	13	2
Kcr_1076	+	1086797	1087786	330	NOL1/NOP2/sun family putative RNA methylase	nop2p		arCOG00973	Sun	J	tRNA or rRNA cytosine-C5-methylase	12	1
Kcr_1077	+	1087783	1088193	137	conserved hypothetical protein			arCOG02285/3	-	S	Uncharacterized conserved protein	0	5
Kcr_1078	+	1088227	1089657	477	inosine-5'-monophosphate dehydrogenase	impdh		arCOG00612	GuaB	F	IMP dehydrogenase/GMP reductase	4	2
Kcr_1079	-	1089658	1090716	353	Polyprenyl synthetase			arCOG01726	IspA	H	Geranylgeranyl pyrophosphate synthase	13	2
Kcr_1080	-	1090806	1092692	629	molybdopterin oxidoreductase			arCOG01491	BisC	C	Anaerobic dehydrogenase	10	2
Kcr_1081	+	1092776	1093009	78	ferredoxin			arCOG00349	Fer	C	Ferredoxin	8	1
Kcr_1082	-	1093012	1095621	870	DEAD/H associated			arCOG00557	Lhr	R	Lhr-like helicase	13	2

Gene	Strand	Start	Stop	Size (a.a.)	Product	Gene Name	COGs	arCOGs	COG_gene	Func	arCOG annotation	# of Cren in arCOGs	#
Kcr_1083	+	1095861	1096211	117	4Fe-4S ferredoxin, iron-sulfur binding			arCOG01500/2	HybA	C	Fe-S-cluster-containing hydrogenase component 1	11	1
Kcr_1084	+	1096213	1097949	579	aldehyde ferredoxin oxidoreductase			arCOG00706	-	C	Aldehyde:ferredoxin oxidoreductase	9	1
Kcr_1085	-	1097946	1098443	166	hypothetical protein			arCOG07435	0	R	HAD superfamily phosphatase	2	0
Kcr_1086	-	1098474	1099208	245	conserved hypothetical protein			arCOG01150	-	R	Predicted ICC-like phosphoesterase	13	2
Kcr_1087	-	1099205	1099633	143	6,7-dimethyl-8-ribityllumazine synthase			arCOG01323/3	RibH	H	Riboflavin synthase beta-chain	11	2
Kcr_1088	+	1099711	1101519	603	DEAD_2			arCOG00770	DinG	K	Rad3-related DNA helicase	12	2
Kcr_1089	-	1101525	1102259	245	Shikimate kinase, archaea			arCOG01025	-	E	Archaeal shikimate kinase	10	2
Kcr_1090	-	1102300	1103175	292	shikimate 5-dehydrogenase	aroE		arCOG01033	AroE	E	Shikimate 5-dehydrogenase	10	2
Kcr_1091	+	1103236	1104156	307	Methionine synthase, vitamin-B12 independent			arCOG01876	MetE	E	Methionine synthase II (cobalamin-independent)	12	2
Kcr_1092	-	1104143	1104745	201	hypothetical protein			arCOG02039/3	Cls	I	Phosphatidylserine/phosphatidylglycerophosphate/cardiolipin synthase or related enzyme	1	8
Kcr_1093	+	1104999	1106840	614	extracellular solute-binding protein, family 5			arCOG01534	DdpA	E	ABC-type dipeptide transport system, periplasmic component	10	1
Kcr_1094	+	1106935	1107927	331	binding-protein-dependent transport systems inner membrane component			arCOG00751	DppB	E	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	13	2
Kcr_1095	+	1107924	1108853	310	binding-protein-dependent transport systems inner membrane component			arCOG00748	DppC	E	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	12	2
Kcr_1096	+	1108858	1109970	371	Oligopeptide/dipeptide ABC transporter, ATP-binding protein-like			arCOG00181	DppD	E	ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	13	1
Kcr_1097	+	1109967	1110983	339	Oligopeptide/dipeptide ABC transporter, ATP-binding protein-like			arCOG00184	AppF	E	ABC-type oligopeptide transport system, ATPase component	13	1
Kcr_1098	+	1111215	1112744	510	ferredoxin			arCOG02035/1	-	R	Uncharacterized metal-binding protein	2	1
Kcr_1099	-	1112915	1113520	202	hypothetical protein			arCOG01426/2	0	S	Uncharacterized conserved protein	5	0
Kcr_1100	+	1113811	1114815	335	3-dehydroquinate synthase			arCOG04353/2	-	E	Predicted alternative 3-dehydroquinate synthase	1	2
Kcr_1101	-	1114818	1115222	135	hypothetical protein								
Kcr_1102	+	1115279	1115797	173	protein of unknown function DUF84			arCOG01221	-	F	Inosine/xanthosine triphosphatase	12	1
Kcr_1103	+	1115802	1116329	176	hypothetical protein								
Kcr_1104	+	1116367	1116684	106	hypothetical protein								
Kcr_1105	+	1116716	1117030	105	hypothetical protein								
Kcr_1106	-	1117091	1117525	145	CBS			arCOG00606	-	R	CBS domain	11	2
Kcr_1107	+	1117835	1118245	137	CopG family transcriptional regulator, nickel-responsive regulator			arCOG01008	NikR	K	Transcriptional regulator, CopG/Arc/MetJ family (DNA-binding and a metal-binding domains)	12	2
Kcr_1108	-	1118228	1119034	269	cobalt transport protein ATP-binding subunit	cbiO		arCOG00202	CbiO	P	ABC-type cobalt transport system, ATPase component	6	2
Kcr_1109	-	1119035	1119769	245	hypothetical protein								
Kcr_1110	-	1119753	1120031	93	hypothetical protein			arCOG03159	CbiM	P	ABC-type Co2+ transport system, permease component	6	2
Kcr_1111	-	1120028	1120636	203	cobalamin (vitamin B12) biosynthesis CbiM protein			arCOG02248	CbiM	P	ABC-type Co2+ transport system, permease component	6	2
Kcr_1112	-	1120745	1121884	380	hydrogenase formation HypD protein			arCOG04428	HypD	O	Hydrogenase maturation factor	6	1
Kcr_1113	-	1121868	1124180	771	[NiFe] hydrogenase maturation protein HypF	hypF		arCOG01187	HypF	O	Hydrogenase maturation factor	6	1
Kcr_1114	+	1124485	1125780	432	heterodisulfide reductase, subunit E, putative			arCOG00332/3	GlpC	C	Membrane associated Fe-S oxidoreductase	6	7
Kcr_1115	+	1125777	1127057	427	4Fe-4S ferredoxin, iron-sulfur binding			arCOG02235/2	HdrA	C	Heterodisulfide reductase, subunit A or related polyferredoxin	0	1
Kcr_1116	+	1127069	1129039	657	4Fe-4S ferredoxin, iron-sulfur binding			arCOG02235	HdrA	C	Heterodisulfide reductase, subunit A or related polyferredoxin	0	1
Kcr_1117	+	1129045	1129470	142	methyl-viologen-reducing hydrogenase, delta subunit			arCOG02475	FrhD	C	Coenzyme F420-reducing hydrogenase, delta subunit	0	1
Kcr_1118	+	1129470	1130429	320	NADH ubiquinone oxidoreductase, 20 kDa subunit			arCOG02472	FrhG	C	Coenzyme F420-reducing hydrogenase, gamma subunit	0	1
Kcr_1119	+	1130426	1131895	490	nickel-dependent hydrogenase, large subunit			arCOG01549	FrhA	C	Coenzyme F420-reducing hydrogenase, alpha subunit	0	1
Kcr_1120	+	1131916	1132374	153	peptidase M52, hydrogen uptake protein			arCOG04429	HyaD	C	Ni,Fe-hydrogenase maturation factor	7	1
Kcr_1121	+	1132417	1132851	145	hydrogenase expression/synthesis, HypA			arCOG04426	HybF	R	Zn finger protein HypA/HybF (possibly regulating hydrogenase expression)	7	1
Kcr_1122	+	1132858	1133619	254	ATPase involved in chromosome partitioning, ParA/MinD family, Mrp homolog			arCOG00585	Mrp	D	Mrp family protein, ATPase	13	2
Kcr_1123	+	1133625	1133837	71	hydrogenase assembly chaperone hypC/hupF	hypC_hupF		arCOG04427	HypC	O	Hydrogenase maturation factor	5	1
Kcr_1124	+	1133846	1134880	345	Hydrogenase expression/formation protein HypE			arCOG00636	HypE	O	Hydrogenase maturation factor	12	2
Kcr_1125	+	1134926	1135714	263	hypothetical protein			arCOG05634	FdhE	O	Uncharacterized protein involved in formate dehydrogenase formation	3	0
Kcr_1126	+	1135921	1136247	109	twin-arginine translocation protein, TatA/E family	tatAE		arCOG02694	TatA	U	Sec-independent protein secretion pathway component	12	1
Kcr_1127	-	1136298	1136660	121	protein of unknown function UPF0153			arCOG02579/2	-	R	Predicted Fe-S-cluster oxidoreductase	9	2
Kcr_1128	+	1136694	1138052	453	peptidase U62, modulator of DNA gyrase			arCOG00321	TldD	R	Predicted Zn-dependent protease or their inactivated homolog	12	2
Kcr_1129	+	1138057	1139373	439	peptidase U62, modulator of DNA gyrase			arCOG00322	TldD	R	Predicted Zn-dependent protease or their inactivated homolog	12	2
Kcr_1130	+	1139370	1140884	505	GMP synthase, C-terminal domain	guaA_Nterm		arCOG00085/1	GuaA	F	GMP synthase, PP-ATPase domain/subunit	10	2
Kcr_1131	-	1140889	1141212	108	translation initiation factor SUI1			arCOG04223	SUI1	J	Translation initiation factor 1 (eIF-1/SUI1)	13	2
Kcr_1132	-	1141515	1141919	135	hypothetical protein								
Kcr_1133	+	1142363	1142608	82	regulatory proteins, AsnC/Lrp			arCOG01580/2	Lrp	K	Transcriptional regulator (Lrp/AsnC family)	13	2
Kcr_1134	+	1142605	1143117	171	Protein of unknown function DUF367			arCOG04733	-	S	Uncharacterized conserved protein	6	2
Kcr_1135	-	1143088	1143903	272	Prephenate dehydratase			arCOG00255	PheA	E	Prephenate dehydratase	9	2
Kcr_1136	+	1144061	1144933	291	Hydrogenase maturation factor-like			arCOG00636	HypE	O	Hydrogenase maturation factor	12	2
Kcr_1137	+	1145322	1145546	75	transcription factor CBF/NF-Y histone			arCOG02144	HHT1	L	Histones H3 and H4	3	2
Kcr_1138	+	1145682	1147130	483	CoA-binding			arCOG01340	-	C	Acyl-CoA synthetase (NDP forming)	11	2
Kcr_1139	-	1147072	1149570	833	S-layer domain-like			arCOG02079/1	-	M	S-layer domain	1	1
Kcr_1140	-	1149571	1150689	373	ABC-2 type transporter			arCOG01463/3	-	V	ABC-type multidrug transport system, permease component	5	1
Kcr_1141	-	1150694	1151623	310	ABC transporter related			arCOG00194	CcmA	V	ABC-type multidrug transport system, ATPase component	13	2
Kcr_1142	+	1151694	1152089	132	hypothetical protein			arCOG04079/3	0	S	Uncharacterized conserved protein	2	4
Kcr_1143	+	1152310	1152744	145	conserved hypothetical protein			arCOG00743	0	K	Predicted transcriptional regulator	2	4
Kcr_1144	+	1152728	1153945	406	protein of unknown function DUF651, N-terminal			arCOG04269	-	S	Uncharacterized conserved protein	13	2
Kcr_1145	-	1153919	1155025	369	geranylgeranyl reductase			arCOG00570	FixC	C	Dehydrogenase (flavoprotein)	13	2
Kcr_1146	+	1155253	1155771	173	conserved hypothetical protein			arCOG04298	-	S	Uncharacterized conserved protein	12	1
Kcr_1147	-	1155863	1156366	168	ribonuclease HII/HIII			arCOG04121/2	RnhB	L	Ribonuclease HII	13	2
Kcr_1148	-	1156503	1157717	405	hypothetical protein			arCOG02080/3	-	M	S-layer domain	1	1
Kcr_1149	+	1157761	1158504	248	ABC transporter related			arCOG00194	CcmA	V	ABC-type multidrug transport system, ATPase component	13	2
Kcr_1150	+	1158497	1159771	425	ABC-2 type transporter			arCOG01462	NatB	C	ABC-type Na+ efflux pump, permease component	9	1
Kcr_1151	-	1159727	1160224	166	NUDIX hydrolase			arCOG01075/1	-	F	NUDIX family hydrolase	12	2
Kcr_1152	-	1160240	1160653	138	hypothetical protein			arCOG03770/1	0	S	Uncharacterized conserved protein	11	0
Kcr_1153	-	1160656	1161792	379	conserved hypothetical protein			arCOG04181	-	S	Uncharacterized protein conserved in archaea	12	2
Kcr_1154	+	1161916	1162560	215	HAD-superfamily hydrolase subfamily IB, PSPase-like			arCOG01158	SerB	E	Phosphoserine phosphatase	7	2
Kcr_1155	+	1162615	1163268	218	regulatory protein, ArsR			arCOG01685/1	ArsR	K	Transcriptional regulator containing HTH domain, ArsR family	0	8
Kcr_1156	-	1163263	1163523	87	hypothetical protein			arCOG07440	0	S	Uncharacterized conserved protein	2	0
Kcr_1157	+	1163785	1164078	98	protein of unknown function DUF343			arCOG04124	-	S	Uncharacterized conserved protein	13	1
Kcr_1158	+	1164093	1164620	176	Appr-1-p processing			arCOG04225	-	R	Predicted phosphatase homologous to the C-terminal domain of histone macroH2A1	12	1
Kcr_1159	-	1164795	1165145	117	hypothetical protein								
Kcr_1160	+	1165357	1166052	232	conserved hypothetical protein			arCOG01879	0	S	Uncharacterized conserved protein	10	4
Kcr_1161	+	1166049	1166681	211	phosphatidate cytidylyltransferase			arCOG01882	SEC59	I	Dolichol kinase	6	0
Kcr_1162	+	1166857	1167984	376	DegT/DnrJ/EryC1/StrS aminotransferase			arCOG00118	WecE	M	Predicted pyridoxal phosphate-dependent enzyme apparently involved in regulation of cell wall biogenesis	8	1
Kcr_1163	-	1167971	1169608	546	CTP synthase			arCOG00063	PyrG	F	CTP synthase (UTP-ammonia lyase)	13	2
Kcr_1164	-	1169689	1169958	90	hypothetical protein								
Kcr_1165	-	1170063	1170788	242	hypothetical protein								
Kcr_1166	-	1170860	1172269	470	cytochrome b/b6-like			arCOG01721	QcrB	C	Cytochrome b subunit of the bc complex	9	6
Kcr_1167	-	1172266	1172769	168	Twin-arginine translocation pathway signal			arCOG01720	QcrA	C	Rieske Fe-S protein	8	4
Kcr_1168	+	1173155	1174873	573	succinate dehydrogenase or fumarate reductase, flavoprotein subunit	sdhA		arCOG00571	SdhA	C	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit	11	1
Kcr_1169	+	1174884	1175300	139	conserved hypothetical protein			arCOG02244/1	SdhC	C	Succinate dehydrogenase/fumarate reductase, cytochrome b subunit	9	7
Kcr_1170	+	1175323	1175625	101	hypothetical protein			arCOG04162	SdhD	C	Succinate dehydrogenase, hydrophobic anchor subunit	4	0
Kcr_1171	+	1175638	1176450	271	succinate dehydrogenase and fumarate reductase iron-sulfur protein	dhsB		arCOG00962	FrdB	C	Succinate dehydrogenase/fumarate reductase, Fe-S protein subunit	12	9
Kcr_1172	+	1176672	1177328	219	Desulfoferrodoxin-like			arCOG02146/1	-	C	Desulfoferrodoxin	3	1
Kcr_1173	+	1177315	1177944	210	HhH-GPD			arCOG04357	-	L	Thermostable 8-oxoguanine DNA glycosylase	3	6
Kcr_1174	+	1177963	1178457	165	conserved hypothetical protein			arCOG07487	-	S	Predicted membrane protein	2	0
Kcr_1175	+	1178646	1179731	362	hydroxymethylglutaryl-CoA synthase, putative	hmgcas		arCOG01767	PksG	I	3-hydroxy-3-methylglutaryl CoA synthase	13	2
Kcr_1176	+	1179742	1180878	379	beta-ketoacyl synthase			arCOG01278	PaaJ	I	Acetyl-CoA acetyltransferase	13	2
Kcr_1177	+	1180881	1181294	138	protein of unknown function DUF35			arCOG01285	-	V	Predicted nucleic-acid-binding protein containing a Zn-ribbon	13	2
Kcr_1178	+	1181474	1182799	442	Glutamyl-tRNA(Gln) amidotransferase, subunit E	asn1		arCOG01924	AnsB	E	L-asparaginase/archaeal Glu-tRNAGln amidotransferase subunit D	13	2
Kcr_1179	+	1182804	1184699	632	aspartyl-tRNA(Asn) amidotransferase, B subunit, putative	gatB_rel		arCOG01719	GatE	J	Archaeal Glu-tRNAGln amidotransferase subunit E (contains GAD domain)	13	2
Kcr_1180	-	1184676	1187162	829	DNA polymerase B region			arCOG00328	PolB	L	DNA polymerase elongation subunit (family B)	13	2

Gene	Strand	Start	Stop	Size (a.a.)	Product	Gene Name	COGs	arCOGs	COG_gene	Func	arCOG annotation	# of Cren in arCOGs	#
Kcr_1181	+	1187232	1188203	324	hypothetical protein								
Kcr_1182	+	1188311	1188778	156	ribosomal protein L22	L22		arCOG04098	RplV	J	Ribosomal protein L22	13	2
Kcr_1183	+	1188775	1189530	252	ribosomal protein S3	rpsC		arCOG04097	RpsC	J	Ribosomal protein S3	13	2
Preliminary coordinates		1189520	1189693	58	ribosomal protein L29			arCOG00785	RpmC	J	Ribosomal protein L29	13	2
Kcr_1184	+	1189762	1189980	73	Ribonuclease P subunit			arCOG00784	POP4	J	RNase P/RNase MRP subunit p29	9	2
Kcr_1185	+	1189977	1190309	111	ribosomal protein S17			arCOG04096	RpsQ	J	Ribosomal protein S17	13	2
Kcr_1186	-	1190278	1191396	373	glycine cleavage system T protein	gcvT		arCOG00756	GcvT	E	Glycine cleavage system T protein (aminomethyltransferase)	6	1
Kcr_1187	+	1191467	1192807	447	FAD linked oxidase-like			arCOG00337	GlcD	C	FAD/FMN-containing dehydrogenase	12	1
Kcr_1188	+	1192817	1193896	360	4Fe-4S ferredoxin, iron-sulfur binding			arCOG00333	GlpC	C	Fe-S oxidoreductase	10	2
Kcr_1189	-	1193874	1195262	463	UDP-glucose/GDP-mannose dehydrogenase			arCOG00252	WecC	M	UDP-N-acetyl-D-mannosaminuronate dehydrogenase	8	1
Kcr_1190	-	1195359	1195562	68	30S ribosomal protein S17e			arCOG01885	RPS17A	J	Ribosomal protein S17E	13	2
Kcr_1191	+	1195764	1196969	402	TOPRIM			arCOG04281	DnaG	L	DNA primase (bacterial type)	13	2
Kcr_1192	+	1196982	1198049	356	putative translation factor pelota	pelota		arCOG01741	PelA	R	Predicted RNA-binding protein	13	2
Kcr_1193	-	1198157	1198867	237	conserved hypothetical protein			arCOG04066	-	R	Predicted nucleotidyltransferase	7	7
Kcr_1194	+	1198871	1199545	225	hypothetical protein			arCOG00042	MesJ	D	tRNA(Ile)-lysidine synthase MesJ	13	2
Kcr_1195	+	1199558	1200997	480	prolyl-tRNA synthetase	proS1		arCOG00402	ProS	J	Prolyl-tRNA synthetase	13	2
Kcr_1196	+	1201115	1201456	114	Ribosomal protein S26E			arCOG04305	RPS26B	J	Ribosomal protein S26	13	0
Kcr_1197	-	1201457	1203637	727	DNA polymerase B region			arCOG00328	PolB	L	DNA polymerase elongation subunit (family B)	13	2
Kcr_1198	+	1203679	1204800	374	thiamine biosynthesis protein			arCOG00038	ThiI	H	Thiamine biosynthesis ATP pyrophosphatase	13	2
Kcr_1199	+	1204993	1207941	983	isoleucyl-tRNA synthetase	ileS		arCOG00807	IleS	J	Isoleucyl-tRNA synthetase	13	2
Kcr_1200	+	1207941	1209200	420	aspartyl-tRNA synthetase	aspSa		arCOG00406	AsnS	J	Aspartyl/asparaginyl-tRNA synthetase	13	2
Kcr_1201	+	1209264	1211675	804	hypothetical protein			arCOG02080/3	-	M	S-layer domain	1	1
Kcr_1202	+	1211666	1212712	349	hypothetical protein			arCOG00427/3	RecJ	L	Single-stranded DNA-specific exonuclease RecJ	3	2
Kcr_1203	+	1212771	1213100	110	hypothetical protein			arCOG02736	-	S	NifX family protein	0	6
Kcr_1204	+	1213185	1213934	250	protein of unknown function DUF549			arCOG03218	-	S	Uncharacterized conserved protein	0	1
Kcr_1205	+	1213974	1214477	168	Protein of unknown function UPF0025			arCOG01141	-	R	Predicted phosphoesterase	9	2
Kcr_1206	-	1214472	1215641	390	aminotransferase, class I and II			arCOG01130	-	E	Aspartate/tyrosine/aromatic aminotransferase	13	2
Kcr_1207	-	1215669	1216673	335	conserved hypothetical protein 374			arCOG00899	-	S	Uncharacterized conserved membrane protein	3	2
Kcr_1208	-	1216684	1217589	302	PfkB			arCOG00014	RbsK	G	Sugar kinase, ribokinase family	13	2
Kcr_1209	-	1217586	1218161	192	Thymidylate kinase			arCOG01891	Tmk	F	Thymidylate kinase	13	2
Kcr_1210	+	1218214	1219119	302	ribose-phosphate pyrophosphokinase	rpppk		arCOG00067	PrsA	F	Phosphoribosylpyrophosphate synthetase	13	2
Kcr_1211	+	1219116	1219493	126	hypothetical protein								
Kcr_1212	+	1219587	1220597	337	DNA repair and recombination protein RadA			arCOG00415	RecA	L	RecA/RadA recombinase	13	2
Kcr_1213	-	1220594	1221229	212	transferase hexapeptide repeat			arCOG01848	WbbJ	R	Acetyltransferase (isoleucine patch superfamily)	8	1
Kcr_1214	+	1221300	1221521	74	hypothetical protein			arCOG00535	ThiS	H	Sulfur transfer protein involved in thiamine biosynthesis	5	2
Kcr_1215	+	1221851	1222837	329	Protein of unknown function UPF0021			arCOG00042	MesJ	D	tRNA(Ile)-lysidine synthase MesJ	13	2
Kcr_1216	-	1222803	1224107	435	phosphonopyruvate decarboxylase-related protein	bcpB		arCOG01696	-	G	Predicted phosphoglycerate mutase, AP superfamily	13	2
Kcr_1217	+	1224160	1224927	256	protein of unknown function DUF75			arCOG00348	-	R	Uncharacterized protein (ATP-grasp superfamily)	1	2
Kcr_1218	+	1224950	1225264	105	transcriptional regulator PadR-like			arCOG00001	-	K	Predicted transcriptional regulator, PadR family	10	1
Kcr_1219	+	1225261	1226277	339	protein of unknown function DUF354			arCOG01395	-	R	Lipid-A-disaccharide synthase related glycosyltransferase	7	1
Kcr_1220	-	1226274	1227626	451	FAD-dependent pyridine nucleotide-disulphide oxidoreductase			arCOG01069	HcaD	R	NAD(FAD)-dependent dehydrogenase	10	2
Kcr_1221	+	1227663	1228418	252	protein of unknown function DUF75			arCOG00348	-	R	Uncharacterized protein (ATP-grasp superfamily)	1	2
Kcr_1222	+	1228429	1229415	329	band 7 protein			arCOG01915/3	HflC	O	Membrane protease subunit, stomatin/prohibitin homolog	12	2
Kcr_1223	+	1229416	1229829	138	protein of unknown function DUF296			arCOG04212	-	R	Predicted DNA-binding protein with PD1-like DNA-binding motif	4	2
Kcr_1224	-	1229826	1230692	289	stationary-phase survival protein SurE	surE		arCOG02303	SurE	R	Predicted acid phosphatase	8	2
Kcr_1225	-	1230682	1231395	238	Glutaredoxin-like			arCOG01218	TrxA	O	Thiol-disulfide isomerase or thioredoxin	12	8
Kcr_1226	+	1231873	1232481	203	2-oxoacid:acceptor oxidoreductase, gamma subunit, pyruvate/2-ketoisovalerate			arCOG01603	PorG	C	Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidoreductase, gamma subunit	11	2
Kcr_1227	+	1232478	1232753	92	2-oxoacid:acceptor oxidoreductase, delta subunit, pyruvate/2-ketoisovalerate			arCOG01605	-	C	Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidoreductase, delta subunit	10	2
Kcr_1228	+	1232794	1233999	402	pyruvate flavodoxin/ferredoxin oxidoreductase-like			arCOG01608	PorA	C	Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidoreductase, alpha subunit	11	2
Kcr_1229	+	1233999	1234979	327	thiamine pyrophosphate enzyme-like TPP-binding			arCOG01601	PorB	C	Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidoreductase, beta subunit	11	2
Kcr_1230	-	1235035	1235274	80	hypothetical protein								
Kcr_1231	-	1235331	1235756	142	hypothetical protein			arCOG06168	DsbB	O	Disulfide bond formation protein DsbB	0	2
Kcr_1232	-	1235928	1236173	82	Protein of unknown function DUF504			arCOG01302	-	S	Uncharacterized protein conserved in archaea	12	1
Kcr_1233	+	1236231	1236938	236	metallophosphoesterase			arCOG01147	Icc	R	Predicted phosphohydrolase	7	1
Kcr_1234	+	1236947	1237279	111	CutA1 divalent ion tolerance protein			arCOG04231	CutA	P	Uncharacterized protein involved in tolerance to divalent cations	10	2
Kcr_1235	+	1237404	1237943	180	UBA/THIF-type NAD/FAD binding fold			arCOG01676/2	ThiF	H	Dinucleotide-utilizing enzyme involved in molybdopterin and thiamine biosynthesis	13	2
Kcr_1236	-	1237944	1239470	509	glycyl-tRNA synthetase	glyS		arCOG00405/3	GRS1	J	Glycyl-tRNA synthetase (class II)	13	2
Kcr_1237	-	1239451	1240026	192	phosphoribosyltransferase			arCOG00040	-	R	Predicted phosphoribosyltransferase	12	1
Kcr_1238	-	1240029	1240319	97	MazG nucleotide pyrophosphohydrolase			arCOG01084	MazG	R	Predicted pyrophosphatase	9	1
Kcr_1239	+	1240492	1241475	328	putative mRNA 3-end processing factor			arCOG00545	YSH1	J	Predicted exonuclease of the beta-lactamase fold involved in RNA processing	13	1
Kcr_1240	-	1241470	1241967	166	regulatory proteins, AsnC/Lrp			arCOG01580	Lrp	K	Transcriptional regulator (Lrp/AsnC family)	13	2
Kcr_1241	+	1242053	1243726	558	membrane protein-like			arCOG04351	-	S	Predicted membrane protein	11	1
Kcr_1242	+	1244037	1244414	126	hypothetical protein								
Kcr_1243	-	1244382	1245635	418	protein of unknown function DUF214			arCOG02312	SalY	V	ABC-type antimicrobial peptide transport system, permease component	10	2
Kcr_1244	-	1245622	1247262	547	hypothetical protein			arCOG04400/1	-	M	S-layer domain	0	1
Kcr_1245	-	1247249	1247974	242	ABC transporter related			arCOG00922	SalX	V	ABC-type antimicrobial peptide transport system, ATPase component	12	2
Kcr_1246	-	1248048	1248431	128	hypothetical protein								
Kcr_1247	+	1248496	1249782	429	histidyl-tRNA synthetase	hisS1		arCOG00404	HisS	J	Histidyl-tRNA synthetase	13	2
Kcr_1248	-	1249730	1250410	227	hypothetical protein			arCOG03727	0	S	Uncharacterized conserved protein	8	2
Kcr_1249	+	1250663	1251946	428	Tryptophan synthase, beta chain-like			arCOG01432	-	R	Predicted alternative tryptophan synthase beta-subunit (paralog of TrpB)	11	1
Kcr_1250	-	1251948	1252880	311	carbamate kinase	arcC		arCOG00863	ArcC	E	Carbamate kinase	9	9
Kcr_1251	-	1252915	1253853	313	ornithine carbamoyltransferase	otc		arCOG00912	ArgF	E	Ornithine carbamoyltransferase	13	2
Kcr_1252	+	1254037	1255242	402	Extracellular ligand-binding receptor			arCOG01020	LivK	E	ABC-type branched-chain amino acid transport system, periplasmic component	5	5
Kcr_1253	+	1255284	1256198	305	inner-membrane translocator			arCOG01270	LivH	E	Branched-chain amino acid ABC-type transport system, permease component	5	4
Kcr_1254	+	1256198	1257151	318	inner-membrane translocator			arCOG01274	LivM	E	ABC-type branched-chain amino acid transport system, permease component	5	5
Kcr_1255	+	1257148	1257888	247	ABC transporter related			arCOG00926	LivG	E	ABC-type branched-chain amino acid transport system, ATPase component	5	4
Kcr_1256	+	1257875	1258582	236	ABC transporter related			arCOG00924	LivF	E	ABC-type branched-chain amino acid transport system, ATPase component	9	6
Kcr_1257	-	1258586	1259683	366	protein of unknown function DUF521			arCOG04278	-	R	Predicted aconitase	10	2
Kcr_1258	-	1260007	1261104	366	Cellulase			arCOG01518	FrvX	G	Cellulase M or related protein	9	2
Kcr_1259	+	1261163	1262146	328	Glutamine-fructose-6-phosphate transaminase (isomerizing)			arCOG00057/2	GlmS	M	Glucosamine 6-phosphate synthetase	13	2
Kcr_1260	+	1262143	1262799	219	hydrolase of the HAD superfamily			arCOG02291	-	R	HAD superfamily hydrolase	11	2
Kcr_1261	+	1262904	1263239	112	Cytidyltransferase-related			arCOG01223/2	-	R	Predicted nucleotidyltransferase	7	2
Kcr_1262	+	1263307	1264428	374	pyruvate flavodoxin/ferredoxin oxidoreductase-like			arCOG01607	PorA	C	Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidoreductase, alpha subunit	0	1
Kcr_1263	+	1264433	1265278	282	thiamine pyrophosphate enzyme-like TPP-binding			arCOG01599	PorB	C	Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidoreductase, beta subunit	13	2
Kcr_1264	+	1265281	1265610	110	4Fe-4S ferredoxin, iron-sulfur binding			arCOG00959/1	-	C	Ferredoxin	0	1
Kcr_1265	+	1265607	1266155	183	2-oxoacid:acceptor oxidoreductase, gamma subunit, pyruvate/2-ketoisovalerate			arCOG01602	PorG	C	Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidoreductase, gamma subunit	11	2
Kcr_1266	+	1266257	1266940	228	Rhomboid-like protein			arCOG01768	GlpG	E	Membrane associated serine protease	12	1
Kcr_1267	+	1266912	1267508	199	class II aldolase/adducin-like			arCOG04226	AraD	G	Ribulose-5-phosphate 4-epimerase or related epimerase/aldolase	9	2
Kcr_1268	-	1267501	1268832	444	UDP-glucose 6-dehydrogenase			arCOG00252	WecC	M	UDP-N-acetyl-D-mannosaminuronate dehydrogenase	8	1
Kcr_1269	+	1268931	1269800	290	Polyprenyl synthetase			arCOG01727	IspA	H	Geranylgeranyl pyrophosphate synthase	11	8
Kcr_1270	-	1269824	1270534	237	hypothetical protein								
Kcr_1271	+	1270879	1271691	271	conserved hypothetical protein			arCOG00238/2	ArsB	P	Na+/H+ antiporter NhaD or related arsenite permease	11	1
Kcr_1272	-	1271676	1272080	135	PilT protein-like			arCOG02730	-	V	PIN domain containing protein	6	5
Kcr_1273	-	1272374	1272736	121	hypothetical protein								
Kcr_1274	-	1272733	1273098	122	hypothetical protein			arCOG04275	0	S	Uncharacterized conserved protein	4	0
Kcr_1275	-	1273125	1273847	241	transcriptional regulator TrmB			arCOG02037	-	K	Sugar-specific transcriptional regulator TrmB	9	2
Kcr_1276	-	1273840	1274148	103	hypothetical protein								
Kcr_1277	-	1274204	1275580	459	Hydroxypyruvate reductase			arCOG04170	GckA	G	Putative glycerate kinase	12	1

Gene	Strand	Start	Stop	Size (a.a.)	Product	Gene Name	COGs	arCOGs	COG_gene	Func	arCOG annotation	# of Cren in arCOGs	#
Kcr_1278	-	1275577	1276812	412	conserved hypothetical protein			arCOG04127	-	R	Predicted GTPase or GTP-binding protein	12	9
Kcr_1279	+	1276983	1277726	248	protein of unknown function DUF72			arCOG04291	-	S	Uncharacterized conserved protein	12	1
Kcr_1280	-	1277714	1278997	428	Actin/actin-like			arCOG05583	-	R	Actin related protein	6	0
Kcr_1281	-	1278981	1279586	202	methylated-DNA--protein-cysteine methyltransferase	ogt		arCOG02724/3	Ada	L	Methylated DNA-protein cysteine methyltransferase	13	2
Kcr_1282	+	1279652	1280698	349	RNA-3'-phosphate cyclase			arCOG04125	RCL1	J	RNA 3'-terminal phosphate cyclase	13	1
Kcr_1283	+	1280688	1282349	554	nucleic acid binding, OB-fold, tRNA/helicase-type			arCOG01510/3	RFA1	L	Single-stranded DNA-binding replication protein A (RPA), large (70 kD) subunit or related ssDNA-binding protein	8	2
Kcr_1284	+	1282330	1282977	216	ERCC4			arCOG00872/2	MPH1	L	ERCC4-like helicase	1	2
Kcr_1285	-	1282968	1284266	433	asparaginyl-tRNA synthetase	asnS		arCOG00407	AsnS	J	Aspartyl/asparaginyl-tRNA synthetase	7	8
Kcr_1286	+	1284349	1285686	446	Carboxylyase-related protein			arCOG01671	UbiD	H	3-polyprenyl-4-hydroxybenzoate decarboxylase or related decarboxylase	13	2
Kcr_1287	+	1285723	1286817	365	Small GTP-binding protein domain			arCOG00353	HflX	R	GTP-binding protein protease modulator	13	1
Kcr_1288	+	1286870	1287421	184	hypothetical protein								
Kcr_1289	-	1287335	1288189	285	Arginase/agmatinase/formiminoglutamase			arCOG01700	SpeB	E	Arginase family enzyme	13	2
Kcr_1290	+	1288458	1288712	85	hypothetical protein								
Kcr_1291	+	1288798	1289067	90	2-oxoacid:acceptor oxidoreductase, delta subunit, pyruvate/2-ketoisovalerate			arCOG01605	-	C	Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidoreductase, delta subunit	10	2
Kcr_1292	+	1289060	1290286	409	pyruvate flavodoxin/ferredoxin oxidoreductase-like			arCOG01608	PorA	C	Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidoreductase, alpha subunit	11	2
Kcr_1293	+	1290283	1291170	296	thiamine pyrophosphate enzyme-like TPP-binding			arCOG01601	PorB	C	Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidoreductase, beta subunit	11	2
Kcr_1294	-	1291115	1292671	519	hypothetical protein		COG3975			E			
Kcr_1295	+	1292735	1294039	435	amidohydrolase			arCOG00695	SsnA	F	Cytosine deaminase or related metal-dependent hydrolase	7	2
Kcr_1296	-	1294019	1294585	189	adenylate cyclase			arCOG01723/2	CyaB	F	Adenylate cyclase, class 2 (thermophilic)	12	2
Kcr_1297	+	1294871	1295224	118	hypothetical protein			arCOG04102		C	Archaeal/vacuolar-type H+-ATPase subunit F	13	2
Kcr_1298	+	1295234	1295530	99	archaeal/vacuolar-type H+-ATPase, subunit H			arCOG03363	NtpF	C	Archaeal/vacuolar-type H+-ATPase subunit H	0	2
Kcr_1299	+	1295523	1296143	207	H+-transporting two-sector ATPase, E subunit			arCOG00869	NtpE	C	Archaeal/vacuolar-type H+-ATPase subunit E	13	2
Kcr_1300	+	1296144	1297931	596	Sodium-transporting two-sector ATPase			arCOG00868	NtpA	C	Archaeal/vacuolar-type H+-ATPase subunit A	13	2
Kcr_1301	+	1297931	1299328	466	Sodium-transporting two-sector ATPase			arCOG00865	NtpB	C	Archaeal/vacuolar-type H+-ATPase subunit B	13	2
Kcr_1302	+	1299333	1299950	206	V-type ATPase, D subunit	vatpD		arCOG04101	NtpD	C	Archaeal/vacuolar-type H+-ATPase subunit D	13	2
Kcr_1303	+	1299950	1300972	341	H+-transporting two-sector ATPase, C (AC39) subunit			arCOG02459	NtpC	C	Archaeal/vacuolar-type H+-ATPase subunit C	10	2
Kcr_1304	-	1301038	1301355	106	H+-transporting two-sector ATPase, C subunit			arCOG02455	AtpK	C	Archaeal/vacuolar-type Na+/H+-ATPase, subunit K	12	2
Kcr_1305	-	1301400	1303853	818	V-type ATPase, 116 kDa subunit			arCOG04138/1	Ntpl	C	Archaeal/vacuolar-type H+-ATPase subunit I	13	2
Kcr_1306	+	1303862	1305223	454	UDP-N-acetylglucosamine 2-epimerase			arCOG01392	WecB	M	UDP-N-acetylglucosamine 2-epimerase	2	1
Kcr_1307	+	1305270	1305899	210	Orotidine 5'-phosphate decarboxylase			arCOG00053/2	SgbH	G	3-hexulose-6-phosphate synthase or related protein	12	1
Kcr_1308	-	1305889	1308765	959	DEAD/DEAH box helicase-like			arCOG00557	Lhr	R	Lhr-like helicase	13	2
Kcr_1309	+	1308913	1310556	548	thermosome			arCOG01257	GroL	O	Chaperonin GroEL (HSP60 family)	13	2
Kcr_1310	+	1310559	1310813	85	RNA polymerase Rpb6			arCOG01268/1	RpoZ	K	DNA-directed RNA polymerase, subunit K/omega	13	2
Kcr_1311	+	1310888	1311169	94	Alba, DNA/RNA-binding protein			arCOG01753	Ssh10b	K	Archaeal DNA-binding protein	13	1
Kcr_1312	+	1311240	1312859	540	lysyl-tRNA synthetase	lysSa		arCOG00485	LysS	J	Lysyl-tRNA synthetase (class I)	5	2
Kcr_1313	+	1312846	1313322	159	hypothetical protein			arCOG05463	0	S	Uncharacterized conserved protein	6	0
Kcr_1314	-	1313312	1313578	89	hypothetical protein								
Kcr_1315	-	1313568	1314371	268	hypothetical protein								
Kcr_1316	-	1314368	1314637	90	hypothetical protein								
Kcr_1317	-	1314897	1315298	134	hypothetical protein								
Kcr_1318	+	1315525	1315788	88	hypothetical protein								
Kcr_1319	-	1315848	1316690	281	hypothetical protein								
Kcr_1320	-	1316725	1318164	480	type II secretion system protein E			arCOG01818/3	VirB11	N	Type IV secretory pathway, VirB11 component, or related ATPase involved in archaeal flagella biosynthesis	4	1
Kcr_1321	-	1318151	1318786	212	hypothetical protein								
Kcr_1322	-	1318783	1320894	704	Protein of unknown function DUF11			arCOG02487/3	0	S	Cell suface protein	3	1
Kcr_1323	+	1321115	1321765	217	Suppressor Mra1			arCOG04122	Mra1	S	Uncharacterized conserved protein	13	1
Kcr_1324	+	1321762	1322067	102	RNAse P, Rpr2/Rpp21 subunit			arCOG04345	RPR2	J	RNase P subunit RPR2	7	2
Kcr_1325	+	1322080	1322565	162	Ribosomal protein S19e			arCOG01344	RPS19A	J	Ribosomal protein S19E (S16A)	13	2
Kcr_1326	+	1322562	1322882	107	DNA-binding TFAR19-related protein			arCOG04179	-	R	DNA-binding protein	13	2
Kcr_1327	+	1323054	1323314	87	Ribosomal protein L31e			arCOG04473/3	RPL31A	J	Ribosomal protein L31E	13	2
Kcr_1328	+	1323323	1323991	223	translation initiation factor eIF-6, putative	eIF-6		arCOG04176	TIF6	J	Translation initiation factor 6 (eIF-6)	13	2
Kcr_1329	+	1324012	1324398	129	Prefoldin alpha-like			arCOG01341/2	GIM5	O	Predicted prefoldin, molecular chaperone implicated in de novo protein folding	13	2
Kcr_1330	+	1324388	1325260	291	GTP-binding signal recognition particle SRP54, G-domain			arCOG01227	FtsY	U	Signal recognition particle GTPase	13	2
Kcr_1331	+	1325281	1326210	310	ornithine carbamoyltransferase	otc		arCOG00912	ArgF	E	Ornithine carbamoyltransferase	13	2
Kcr_1332	+	1332111	1332356	82	hypothetical protein								
Kcr_1333	+	1332794	1333387	198	conserved hypothetical protein			arCOG01423	-	S	Uncharacterized conserved protein containing a coiled-coil domain	4	0
Kcr_1334	-	1333773	1334753	327	hypothetical protein								
Kcr_1335	-	1335575	1335778	68	4-oxalocrotonate tautomerase			arCOG02240	-	R	4-oxalocrotonate tautomerase related protein	1	9
Kcr_1336	-	1335822	1335983	54	Ribosomal protein S27a			arCOG04183	RPS31	J	Ribosomal protein S27AE	13	2
Kcr_1337	-	1335980	1336267	96	30S ribosomal protein S24e			arCOG04182	RPS24A	J	Ribosomal protein S24E	12	2
Kcr_1338	-	1336338	1336529	64	DNA-directed RNA polymerase subunit E, RpoE2			arCOG04077	-	K	DNA-directed RNA polymerase, subunit E"	13	2
Kcr_1339	-	1336526	1337101	192	RNA binding S1			arCOG00675	RPB7	K	DNA-directed RNA polymerase, subunit E'	13	2
Kcr_1340	-	1337137	1337523	129	Nucleotide binding protein, PINc			arCOG04312	-	R	Uncharacterized protein of PilT N-term./Vapc superfamily	13	2
Kcr_1341	-	1337797	1339020	408	peptide chain release factor eRF/aRF, subunit 1	eRF		arCOG01742	eRF1	J	Peptide chain release factor 1 (eRF1)	13	2
Kcr_1342	+	1339092	1340261	390	N-acetylglucosamine-1-phosphate uridyltransferase -like			arCOG05599/3	GCD1	M	Nucleoside-diphosphate-sugar pyrophosphorylase involved in lipopolysaccharide biosynthesis/translation initiation factor 2B, gamma/epsilon subunit	5	0
Kcr_1343	+	1341178	1342371	398	cell division protein FtsZ	ftsZ		arCOG02201	FtsZ	D	Cell division GTPase	1	2
Kcr_1344	+	1342566	1343045	160	transcription antitermination protein NusG			arCOG01920	NusG	K	Transcription antiterminator NusG	13	2
Kcr_1345	+	1343038	1343517	160	ribosomal protein L11			arCOG04372	RplK	J	Ribosomal protein L11	13	2
Kcr_1346	+	1343519	1344199	227	ribosomal protein L1			arCOG04289	RplA	J	Ribosomal protein L1	13	2
Kcr_1347	+	1344202	1345068	289	ribosomal protein L10			arCOG04288	RplJ	J	Ribosomal protein L10	13	2
Kcr_1348	-	1345065	1346201	379	tyrosyl-tRNA synthetase	tyrS		arCOG01886	TyrS	J	Tyrosyl-tRNA synthetase	13	2
Kcr_1349	-	1346306	1346791	162	protein of unknown function DUF82			arCOG04290	-	R	PIN-domain and Zn ribbon	12	1
Kcr_1350	-	1347079	1347957	293	NAD(+) kinase			arCOG01348	nadF	H	NAD kinase	13	2
Kcr_1351	+	1348013	1348510	166	hypothetical protein								
Kcr_1352	+	1348510	1350621	704	MCM			arCOG00439	MCM2	L	Predicted ATPase involved in replication control, Cdc46/Mcm family	13	2
Kcr_1353	+	1350631	1352718	696	DEAD/DEAH box helicase-like			arCOG00553	-	L	Replicative SFII helicase	12	2
Kcr_1354	+	1352696	1354066	457	adenylosuccinate lyase	purB		arCOG01747	PurB	F	Adenylosuccinate lyase	9	2
Kcr_1355	+	1354083	1355381	433	beta-lactamase-like			arCOG00541	YSH1	J	Predicted exonuclease of the beta-lactamase fold involved in RNA processing	13	1
Kcr_1356	+	1355451	1356095	215	GTP-binding protein, HSR1-related			arCOG00354	-	R	GTPase SAR1 or related small G protein	0	2
Kcr_1357	+	1356105	1356419	105	Rubredoxin-type Fe(Cys)4 protein			arCOG04417	-	R	Zn-ribbon containing protein	0	1
Kcr_1358	-	1356385	1357167	261	Methyltransferase type 12			arCOG00978	GCD14	J	tRNA(1-methyladenosine) methyltransferase or related methyltransferase	13	2
Kcr_1359	-	1357175	1358317	381	geranylgeranyl reductase			arCOG00570	FixC	C	Dehydrogenase (flavoprotein)	13	2
Kcr_1360	+	1358507	1358740	78	hypothetical protein								
Kcr_1361	+	1358771	1360543	591	tRNA-guanine transglycosylases, various specificities	tgt		arCOG00989	Tgt	J	Queuine/archaeosine tRNA-ribosyltransferase	13	2
Kcr_1362	-	1360506	1361009	168	Protein of unknown function DUF359			arCOG04076	-	S	Uncharacterized protein conserved in archaea	13	2
Kcr_1363	+	1361060	1362565	502	Radical SAM			arCOG01356	-	C	Radical SAM superfamily enzyme	9	2
Kcr_1364	+	1362820	1364739	640	hypothetical protein		COG3485			Q			
Kcr_1365	+	1365033	1365818	262	protein of unknown function DUF92, transmembrane			arCOG02245	-	S	Predicted membrane protein	6	2
Kcr_1366	+	1365815	1366546	244	glycosyl transferase, family 2			arCOG01386	-	M	Glycosyltransferase	0	4
Kcr_1367	-	1366705	1367241	179	Brix			arCOG03247	IMP4	J	Predicted exosome subunit/U3 small nucleolar ribonucleoprotein (snoRNP) component, contains IMP4 domain	12	1
Kcr_1368	-	1367238	1367486	83	Ribosomal protein L37AE/L43A-like			arCOG04208/3	RPL43A	J	Ribosomal protein L37AE/L43A	13	2
Kcr_1369	-	1367488	1368294	269	3' exoribonuclease			arCOG01574	-	J	RNase PH-related exoribonuclease	13	1
Kcr_1370	-	1368294	1369040	249	Exosome complex exonuclease 1			arCOG01575	Rph	J	Ribonuclease PH	13	1
Kcr_1371	-	1369045	1369731	229	KH, type 1			arCOG00678	RRP4	J	RNA-binding protein Rrp4 or related protein (contain S1 domain and KH domain)	13	1
Kcr_1372	-	1369724	1370401	226	Shwachman-Bodian-Diamond syndrome proteins			arCOG04187	-	J	Predicted exosome subunit	13	2
Kcr_1373	-	1370407	1371126	240	Proteasome endopeptidase complex			arCOG00971	PRE1	O	20S proteasome, alpha and beta subunit	13	2
Kcr_1374	-	1371417	1371743	109	hypothetical protein								
Kcr_1375	-	1371721	1372362	214	hypothetical protein			arCOG00307	RPP1	J	RNase P/RNase MRP subunit p30	7	2

Gene	Strand	Start	Stop	Size (a.a.)	Product	Gene Name	COGs	arCOGs	COG_gene	Func	arCOG annotation	# of Cren in arCOGs	#
Kcr_1376	-	1372344	1372769	142	Protein of unknown function DUF54			arCOG01042	-	J	Predicted exosome subunit	13	2
Kcr_1377	-	1372766	1373353	196	Ribosomal protein L15e			arCOG04209	RPL15A	J	Ribosomal protein L15E	13	2
Kcr_1378	+	1373415	1374608	398	GTPase of unknown function-like			arCOG00357	-	J	Predicted GTPase, probable translation factor	13	2
Kcr_1379	-	1374601	1375647	349	Nucleotidyl transferase			arCOG00666	GCD1	M	Nucleoside-diphosphate-sugar pyrophosphorylase involved in lipopolysaccharide biosynthesis/translation initiation factor 2B, gamma/epsilon subunit	12	2
Kcr_1380	-	1375619	1376251	211	ribosomal protein S3Ae			arCOG04186	RPS1A	J	Ribosomal protein S3AE	13	2
Kcr_1381	+	1376306	1377127	274	ABC transporter related			arCOG00188/2	CbiO	P	ABC-type cobalt transport system, ATPase component	9	1
Kcr_1382	+	1377132	1377929	266	ABC transporter related			arCOG00202	CbiO	P	ABC-type cobalt transport system, ATPase component	6	2
Kcr_1383	+	1377922	1378665	248	cobalt transport protein			arCOG02250	CbiQ	P	ABC-type cobalt transport system, permease component CbiQ or related transporter	8	2
Kcr_1384	+	1378915	1380105	397	ATPase			arCOG00467	CDC6	L	Cdc6-related protein, AAA superfamily ATPase	13	2
Kcr_1385	-	1380117	1381508	464	DNA-directed DNA polymerase			arCOG04455	HYS2	L	Archaeal DNA polymerase II, small subunit/DNA polymerase delta, subunit B	1	2
Kcr_1386	+	1381629	1382261	211	Excisionase/Xis, DNA-binding			arCOG03164	-	L	Predicted site-specific integrase-resolvase	6	8
Kcr_1387	+	1382251	1383414	388	transposase, IS605 OrfB			arCOG00679	-	L	Transposase	9	1
Kcr_1388	-	1383503	1384258	252	protein of unknown function, ATP binding			arCOG01225	-	R	GTPase SAR1 or related small G protein	13	8
Kcr_1389	-	1384407	1386734	776	aminoacyl-tRNA synthetase, class Ia			arCOG00808	ValS	J	Valyl-tRNA synthetase	13	2
Kcr_1390	-	1386780	1387382	201	protein of unknown function DUF99			arCOG00928	-	L	Endonuclease V homolog	12	2
Kcr_1391	+	1387415	1388131	239	geranylgeranylglyceryl phosphate synthase	gggp		arCOG01085	-	R	Predicted phosphate-binding enzyme, TIM-barrel fold	13	2
Kcr_1392	-	1388470	1389606	379	succinyl-CoA synthetase, beta subunit	sucCoAb		arCOG01337	SucC	C	Succinyl-CoA synthetase, beta subunit	13	1
Kcr_1393	+	1389672	1390547	292	succinyl-CoA synthetase, alpha subunit	sucCoAa		arCOG01339	SucD	C	Succinyl-CoA synthetase, alpha subunit	13	1
Kcr_1394	-	1390557	1391852	432	hypothetical protein								
Kcr_1395	+	1391957	1392874	306	hypothetical protein								
Kcr_1396	-	1392980	1393216	79	hypothetical protein		COG1349			K			
Kcr_1397	+	1393280	1393444	55	Ribosomal protein L40e			arCOG04049	RPL40A	J	Ribosomal protein L40E	13	2
Kcr_1398	+	1393656	1394513	286	Putative cell wall-binding domain-like			arCOG00395/2	-	K	Predicted membrane-associated trancriptional regulator	0	4
Kcr_1399	+	1394547	1394753	69	Protein of unknown function DUF1610			arCOG01989/1	-	J	Predicted Zn-ribbon RNA-binding protein with a function in translation	13	2
Kcr_1400	+	1394755	1395033	93	Elongation factor 1, beta/beta'/delta chain			arCOG01988	EFB1	J	Translation elongation factor EF-1beta	13	2
Kcr_1401	+	1395033	1395281	83	hypothetical protein								
Kcr_1402	-	1395282	1396310	343	XPG I			arCOG04050	Exo	L	5'-3' exonuclease (including N-terminal domain of PolI)	13	2
Kcr_1403	+	1396363	1397427	355	glycosyl transferase, group 1			arCOG01403	RfaG	M	Glycosyltransferase	13	2
Kcr_1404	+	1397494	1397850	119	Aminoacyl-tRNA hydrolase			arCOG04228	pth2	J	Peptidyl-tRNA hydrolase	13	2
Kcr_1405	-	1398278	1398823	182	Putative methylase			arCOG00109	HemK	J	Methylase of polypeptide chain release factors	13	2
Kcr_1406	-	1398820	1399551	244	ribosomal RNA adenine methylase transferase			arCOG04131	KsgA	J	Dimethyladenosine transferase (rRNA methylation)	13	2
Kcr_1407	+	1399592	1401049	486	protein of unknown function UPF0027			arCOG04246	RtcB	R	RtcB family protein, predicted to be implicated in RNA processing	13	2
Kcr_1408	+	1401054	1401740	229	FKBP-type peptidyl-prolyl cis-trans isomerases			arCOG00980	SlpA	O	FKBP-type peptidyl-prolyl cis-trans isomerase 2	13	2
Kcr_1409	-	1401722	1402135	138	Protein of unknown function UPF0179			arCOG04477	-	S	Uncharacterized protein conserved in archaea	5	2
Kcr_1410	+	1402208	1402828	207	Proteasome endopeptidase complex			arCOG00970	PRE1	O	20S proteasome, alpha and beta subunit	13	2
Kcr_1411	+	1402825	1404699	625	beta-lactamase-like			arCOG00543	-	R	Predicted metal-dependent RNase, consists of a metallo-beta-lactamase domain and an RNA-binding KH domain	13	2
Kcr_1412	-	1404696	1405088	131	hypothetical protein			arCOG02263	-	S	Uncharacterized conserved protein	1	2
Kcr_1413	+	1405156	1405398	81	Like-Sm ribonucleoprotein, core			arCOG00998	LSM1	K	Small nuclear ribonucleoprotein (snRNP) homolog	13	2
Kcr_1414	+	1405413	1405556	48	Ribosomal protein L37e			arCOG04126	RPL37A	J	Ribosomal protein L37E	12	2
Kcr_1415	+	1405982	1407331	450	aminotransferase class-III			arCOG00915	GabT	E	4-aminobutyrate aminotransferase or related aminotransferase	12	1
Kcr_1416	+	1407350	1407814	155	hypothetical protein			arCOG02197	-	S	Uncharacterized conserved protein	1	2
Kcr_1417	+	1407849	1408169	107	Prefoldin beta-like			arCOG01342	GimC	O	Prefoldin, chaperonin cofactor	13	2
Kcr_1418	-	1408153	1409490	446	phosphoesterase, RecJ-like			arCOG00427	RecJ	L	Single-stranded DNA-specific exonuclease RecJ	3	2
Kcr_1419	-	1409492	1409947	152	Ribosomal S13/S15-like			arCOG04185	RpsO	J	Ribosomal protein S15P	13	2
Kcr_1420	-	1410136	1410690	185	Ham1-like protein			arCOG04184	-	F	Xanthosine triphosphate pyrophosphatase	13	2
Kcr_1421	-	1410687	1411271	195	protein of unknown function RIO1			arCOG01185	-	T	Mn2+-dependent serine/threonine protein kinase	13	2
Kcr_1422	-	1411262	1411927	222	protein-L-isoaspartate O-methyltransferase	pimt		arCOG00976	Pcm	O	Protein-L-isoaspartate carboxylmethyltransferase	11	2
Kcr_1423	+	1411942	1412349	136	Protein of unknown function DUF371			arCOG04171	-	S	Uncharacterized protein conserved in archaea	12	1
Kcr_1424	-	1412346	1413425	360	Pseudouridylate synthase			arCOG04252	-	J	tRNA pseudouridine synthase D	13	2
Kcr_1425	+	1414184	1414555	124	Sjogrens syndrome scleroderma autoantigen 1			arCOG00578	-	R	Uncharacterized Zn-finger containing protein	13	2
Kcr_1426	-	1414512	1415027	172	hypothetical protein								
Kcr_1427	-	1415021	1416136	372	metal-dependent phosphohydrolase, HD subdomain			arCOG04430	-	R	HD superfamily phosphohydrolase	13	2
Kcr_1428	-	1416199	1416621	141	glycine cleavage system H protein	gcvH		arCOG01303	GcvH	E	Glycine cleavage system H protein (lipoate-binding)	6	1
Kcr_1429	-	1416688	1417182	165	conserved hypothetical protein								
Kcr_1430	-	1417257	1417508	84	conserved hypothetical protein			arCOG04483/3	-	S	Uncharacterized protein containing a ferredoxin domain	0	9
Kcr_1431	-	1417687	1418040	118	ferredoxin domain containing protein			arCOG04483/2	-	S	Uncharacterized protein containing a ferredoxin domain	0	9
Kcr_1432	+	1418328	1418612	95	Protein of unknown function UPF0147			arCOG04308	-	S	Uncharacterized protein conserved in archaea	9	2
Kcr_1433	-	1418602	1419339	246	hypothetical protein								
Kcr_1434	+	1419756	1420163	136	hypothetical protein								
Kcr_1435	-	1420294	1420524	77	hypothetical protein								
Kcr_1436	-	1420614	1421510	299	Rad51-like			arCOG00415	RecA	L	RecA/RadA recombinase	13	2
Kcr_1437	-	1421510	1421986	159	protein of unknown function DUF46			arCOG04106	CdsA	I	CDP-diglyceride synthetase	12	2
Kcr_1438	-	1421986	1422381	132	hypothetical protein								
Kcr_1439	-	1422386	1424089	568	DEAD_2			arCOG00770	DinG	K	Rad3-related DNA helicase	12	2
Kcr_1440	-	1424148	1425308	387	cell division protein FtsZ	ftsZ		arCOG02201	FtsZ	D	Cell division GTPase	1	2
Kcr_1441	-	1425413	1425658	82	CopG-like DNA-binding			arCOG01008/3	NikR	K	Transcriptional regulator, CopG/Arc/MetJ family (DNA-binding and a metal-binding domains)	12	2
Kcr_1442	+	1426005	1427294	430	Glycine dehydrogenase (decarboxylating)			arCOG00077	GcvP	E	Glycine cleavage system protein P (pyridoxal-binding), N-terminal domain	5	1
Kcr_1443	+	1427291	1428829	513	glycine cleavage system P-protein			arCOG00076	GcvP	E	Glycine cleavage system protein P (pyridoxal-binding), C-terminal domain	5	1
Kcr_1444	-	1428776	1429684	303	glycosyl transferase, family 4			arCOG03199	Rfe	M	UDP-N-acetylmuramyl pentapeptide phosphotransferase/UDP-N-acetylglucosamine-1-phosphate transferase	13	1
Kcr_1445	+	1429721	1430698	326	glycosyl transferase, family 4			arCOG03199	Rfe	M	UDP-N-acetylmuramyl pentapeptide phosphotransferase/UDP-N-acetylglucosamine-1-phosphate transferase	13	1
Kcr_1446	+	1431230	1431853	208	nucleic acid binding, OB-fold, tRNA/helicase-type			arCOG02258	-	L	RPA family protein, a subunit of RPA complex in P.uriosus	0	2
Kcr_1447	+	1431858	1432280	141	Translation initiation factor IF5			arCOG01640	GCD7	J	Translation initiation factor 2, beta subunit (eIF-2beta)/eIF-5 N-terminal domain	13	2
Kcr_1448	+	1432290	1432640	117	S1, IF1 type			arCOG01179	IntA	J	Translation initiation factor 1 (IF-1)	13	2
Kcr_1449	+	1432644	1433447	268	protein of unknown function RIO1			arCOG01180	RIO1	T	Serine/threonine protein kinase involved in cell cycle control	13	2
Kcr_1450	+	1433593	1434006	138	KH, type 1			arCOG04150/2	-	R	Predicted RNA-binding protein (contains KH domains)	13	2
Kcr_1451	+	1434023	1435561	513	DNA topoisomerase VI, B subunit	top6b		arCOG01165	-	L	DNA topoisomerase VI, subunit B	13	2
Kcr_1452	+	1435542	1436648	369	DNA topoisomerase VI subunit A (type II DNA topoisomerase)			arCOG04143	-	L	DNA topoisomerase VI, subunit A	13	2
Kcr_1453	-	1437039	1438151	371	Uncharacterized membrane-associated protein/domain-like			arCOG00374/3	-	K	Predicted membrane-associated trancriptional regulator	10	4
Kcr_1454	+	1438236	1438655	140	TOPRIM			arCOG01486	-	L	Ribonuclease M5 (contains TOPRIM domain)	0	1
Kcr_1455	+	1438652	1439101	150	Protein of unknown function DUF54			arCOG01043	-	S	Uncharacterized protein conserved in archaea	13	1
Kcr_1456	-	1439098	1440003	302	Ribonuclease Z			arCOG00501	ElaC	R	Metal-dependent hydrolase of the beta-lactamase superfamily	13	2
Kcr_1457	-	1440047	1440970	308	putative RNA methylase			arCOG00047	-	L	Predicted DNA modification methylase	12	2
Kcr_1458	-	1441038	1442078	347	Small GTP-binding protein domain			arCOG00352	-	R	Predicted GTPase	12	2
Kcr_1459	-	1442083	1443213	377	peptidase M50			arCOG04064	-	M	Predicted membrane-associated Zn-dependent protease	13	2
Kcr_1460	-	1443217	1444320	368	peptidase M20			arCOG01107	ArgE	E	Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase or related deacylase	13	2
Kcr_1461	-	1444332	1444928	199	Proteasome endopeptidase complex			arCOG00970	PRE1	O	20S proteasome, alpha and beta subunit	13	2
Kcr_1462	-	1445046	1446680	545	hypothetical protein								
Kcr_1463	-	1446685	1447026	114	hypothetical protein								
Kcr_1464	+	1447389	1448459	357	Transcription factor TFIIB, cyclin-related			arCOG01981	SUA7	K	Transcription initiation factor TFIIB, Brf1 subunit/Transcription initiation factor TFIIB	13	2
Kcr_1465	+	1448647	1449195	183	peptidase S26B, signal peptidase			arCOG01739/1	LepB	U	Signal peptidase I	13	2
Kcr_1466	+	1449305	1449940	212	TATA-box binding			arCOG01764	SPT15	K	TATA-box binding protein (TBP), component of TFIID and TFIIIB	13	2
Kcr_1467	+	1449987	1451084	366	protein of unknown function DUF100			arCOG04180	-	G	Archaeal fructose 1,6-bisphosphatase	13	1
Kcr_1468	-	1451071	1451757	229	triosephosphate isomerase			arCOG01087	TpiA	G	Triosephosphate isomerase	13	2
Kcr_1469	+	1451803	1452411	203	AMMECR1			arCOG01336	AMMECR1	S	Uncharacterized conserved protein	13	2
Kcr_1470	+	1452412	1453572	387	glycosyltransferase 28-like			arCOG01393	-	G	Glycosyl transferase, related to UDP-glucuronosyltransferase	1	1
Kcr_1471	-	1453795	1454781	329	von Willebrand factor, type A			arCOG02902/3	-	R	Uncharacterized protein containing a von Willebrand factor type A (vWA) domain	2	7
Kcr_1472	-	1454778	1455671	298	ATPase associated with various cellular activities, AAA_5			arCOG00441	-	R	MoxR-like ATPase	8	9
Kcr_1473	-	1455757	1456338	194	Cytidylate kinase, putative			arCOG01037	Cmk	F	Cytidylate kinase	13	2

Gene	Strand	Start	Stop	Size (a.a.)	Product	Gene Name	COGs	arCOGs	COG_gene	Func	arCOG annotation	# of Cren in arCOGs	#
Kcr_1474	-	1456349	1457275	309	metalloendopeptidase, putative, glycoprotease family	gcp		arCOG01183	Kae1/QRI7	L	Iron containing AP-endonuclease	13	2
Kcr_1475	-	1457308	1457937	210	Sua5/YciO/YrdC/YwlC			arCOG01952/2	SUA5	J	Putative translation factor (SUA5)	13	2
Kcr_1476	+	1457996	1458901	302	methionine aminopeptidase, type II	map2		arCOG01001	Map	J	Methionine aminopeptidase	13	2
Kcr_1477	+	1458898	1459308	137	Archaeal holliday junction resolvase			arCOG00919	-	L	Holliday junction resolvase - archaeal type	12	1
Kcr_1478	-	1459298	1460374	359	tRNA intron endonuclease	endA		arCOG01701/1	SEN2	J	tRNA splicing endonuclease	13	2
Kcr_1479	-	1460543	1461478	312	oxidoreductase-like			arCOG01622	MviM	R	Predicted dehydrogenase	9	2
Kcr_1480	+	1461533	1462378	282	UbiA prenyltransferase			arCOG00476	UbiA	H	4-hydroxybenzoate polyprenyltransferase or related prenyltransferase	12	2
Kcr_1481	-	1462365	1463141	259	CBS			arCOG00600	-	R	CBS domain	12	2
Kcr_1482	-	1463131	1463955	275	CBS			arCOG00600	-	R	CBS domain	12	2
Kcr_1483	-	1463967	1464452	162	CBS			arCOG00606	-	R	CBS domain	11	2
Kcr_1484	-	1464458	1466332	625	protein of unknown function DUF814			arCOG01695	-	K	Predicted RNA-binding protein homologous to eukaryotic snRNP	13	2
Kcr_1485	+	1466460	1467311	284	hypothetical protein								
Kcr_1486	+	1467314	1467724	137	protein of unknown function DUF101			arCOG04055	-	R	SHS2 domain protein implicated in nucleic acid metabolism	13	2
Kcr_1487	+	1467878	1468876	333	ribosomal protein L3			arCOG04070	RplC	J	Ribosomal protein L3	13	2
Kcr_1488	+	1468873	1469670	266	ribosomal protein L4/L1e			arCOG04071	RplD	J	Ribosomal protein L4	13	2
Kcr_1489	+	1469663	1469938	92	Ribosomal protein L25/L23			arCOG04072	RplW	J	Ribosomal protein L23	13	2
Kcr_1490	+	1469943	1470680	246	ribosomal protein L2			arCOG04067	RplB	J	Ribosomal protein L2	13	2
Kcr_1491	+	1470711	1471109	133	ribosomal protein S19	rpsS		arCOG04099	RpsS	J	Ribosomal protein S19	13	2
Kcr_1492	-	1471212	1472459	416	conserved hypothetical protein			arCOG02047	-	S	Uncharacterized conserved protein	0	6
Kcr_1493	-	1472461	1473411	317	4-hydroxythreonine-4-phosphate dehydrogenase	pdxA	COG1995			H			
Kcr_1494	-	1473458	1474513	352	conserved hypothetical protein		COG3395			S			
Kcr_1495	-	1474569	1474880	104	ribosomal protein S10			arCOG01758	RpsJ	J	Ribosomal protein S10	13	2
Kcr_1496	-	1474909	1476177	423	translation elongation factor EF-1, subunit alpha	EF-1_alpha		arCOG01561	TEF1	J	Translation elongation factor EF-1alpha (GTPase)	13	2
Kcr_1497	+	1476594	1476974	127	hypothetical protein								
Kcr_1498	-	1476978	1478318	447	phosphoglucomutase/phosphomannomutase alpha/beta/alpha domain I			arCOG00767	{ManB}	G	Phosphomannomutase	13	2
Kcr_1499	-	1478315	1480300	662	Oligopeptide transporter OPT superfamily		COG1297			E			
Kcr_1500	-	1480309	1481061	251	hypothetical protein			arCOG00493/3	GapA	G	Glyceraldehyde-3-phosphate dehydrogenase/erythrose-4-phosphate dehydrogenase	13	2
Kcr_1501	-	1481058	1482266	403	Phosphoglycerate kinase			arCOG00496	Pgk	G	3-phosphoglycerate kinase	13	2
Kcr_1502	-	1482356	1483168	271	purine phosphorylase, family 2			arCOG01327	Pnp	F	Purine nucleoside phosphorylase	13	2
Kcr_1503	+	1483266	1484930	555	thermosome			arCOG01257	GroL	O	Chaperonin GroEL (HSP60 family)	13	2
Kcr_1504	+	1484946	1485830	295	hypothetical protein			arCOG03414/3	0	S	Uncharacterized conserved protein	0	6
Kcr_1505	+	1485841	1487133	431	hypothetical protein			arCOG03206	-	S	Predicted archaeal membrane protein	0	8
Kcr_1506	+	1487130	1488311	394	hypothetical protein								
Kcr_1507	+	1488311	1488688	126	ribosomal protein S8E			arCOG04154	RPS8A	J	Ribosomal protein S8E	13	2
Kcr_1508	+	1488678	1488992	105	Ribonucleoprotein complex SRP, Srp19 component			arCOG01217	SEC65	U	Signal recognition particle 19 kDa protein	12	2
Kcr_1509	-	1489166	1489753	196	protein of unknown function DUF265			arCOG01034	-	F	Predicted nucleotide kinase	12	1
Kcr_1510	+	1489754	1490875	374	tRNA (guanine-N(2)-)-methyltransferase			arCOG01219	TRM1	J	N2,N2-dimethylguanosine tRNA methyltransferase	13	2
Kcr_1511	-	1490862	1491545	228	fibrillarin			arCOG00078	NOP1	J	Fibrillarin-like rRNA methylase	13	2
Kcr_1512	-	1491526	1492674	383	Pre-mRNA processing ribonucleoprotein, binding region			arCOG01923	SIK1	J	Protein implicated in ribosomal biogenesis, Nop56p homolog	13	2
Kcr_1513	-	1492684	1492851	56	30S ribosomal protein S30e			arCOG04293	-	J	Ribosomal protein S30	13	0
Kcr_1514	-	1492969	1494948	660	Protein of unknown function DUF460			arCOG04219	-	R	Possible nuclease of RNase H fold, RuvC/YqgF family	12	1
Kcr_1515	+	1495004	1496287	428	hypothetical protein								
Kcr_1516	+	1496295	1498088	598	glucosamine--fructose-6-phosphate aminotransferase, isomerizing	glmS		arCOG00057	GlmS	M	Glucosamine 6-phosphate synthetase	13	2
Kcr_1517	-	1498089	1498325	79	Like-Sm ribonucleoprotein, core			arCOG00998	LSM1	K	Small nuclear ribonucleoprotein (snRNP) homolog	13	2
Kcr_1518	-	1498363	1498989	209	ribosomal protein S7			arCOG04254	RpsG	J	Ribosomal protein S7	13	2
Kcr_1519	-	1498995	1499453	153	ribosomal protein S23 (S12)	S23_S12		arCOG04255	RpsL	J	Ribosomal protein S12	13	2
Kcr_1520	-	1499440	1499916	159	NusA family KH domain protein	nusA		arCOG01760	NusA	K	Transcription elongation factor	13	2
Kcr_1521	-	1499913	1500236	108	ribosomal protein L7Ae/L30e/S12e/Gadd45			arCOG01752	RPL30	J	Ribosomal protein L30E	13	2
Kcr_1522	-	1500229	1504203	1325	RNA polymerase, alpha subunit			arCOG04256/1	RpoC	K	DNA-directed RNA polymerase subunit A"	13	2
Kcr_1523	-	1504214	1507579	1122	DNA-directed RNA polymerase			arCOG01762	RpoB	K	DNA-directed RNA polymerase subunit B	13	2
Kcr_1524	-	1507583	1507795	71	RNA polymerase subunit, RPB5			arCOG04258	RPB5	K	DNA-directed RNA polymerase, subunit H, RpoH/RPB5	13	2
Kcr_1525	+	1508189	1508782	198	2-oxoacid:acceptor oxidoreductase, gamma subunit, pyruvate/2-ketoisovalerate			arCOG01603	PorG	C	Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidoreductase, gamma subunit	11	2
Kcr_1526	-	1508773	1509012	80	conserved hypothetical protein			arCOG04136/2	-	R	Uncharacterized Zn ribbon-containing protein	12	0
Kcr_1527	+	1509103	1512450	1116	DNA polymerase II, large subunit DP2	polC		arCOG04447	-	L	Archaeal DNA polymerase II, large subunit	1	2
Kcr_1528	+	1512459	1512779	107	nucleic acid binding, OB-fold, tRNA/helicase-type			arCOG05741	0	R	Nucleic acid binding, OB-fold	0	4
Kcr_1529	+	1512769	1513521	251	proliferating cell nuclear antigen (pcna)	pcna		arCOG00488	DnaN	L	DNA polymerase sliding clamp subunit (PCNA homolog)	13	2
Kcr_1530	+	1513502	1514389	296	DNA primase, large subunit			arCOG03013	PRI2	L	Eukaryotic-type DNA primase, large subunit	13	2
Kcr_1531	+	1514368	1515486	373	DNA primase, small subunit			arCOG04110	PRI1	L	Eukaryotic-type DNA primase, catalytic (small) subunit	13	2
Kcr_1532	+	1515476	1515976	167	hypothetical protein			arCOG00551	-	L	DNA replication initiation complex subunit, GINS family	13	2
Kcr_1533	+	1515954	1516232	93	Ribosomal protein L44E			arCOG04109	RPL42A	J	Ribosomal protein L44E	13	2
Kcr_1534	+	1516237	1516437	67	Ribosomal protein S27E			arCOG04108	RPS27A	J	Ribosomal protein S27E	13	2
Kcr_1535	+	1516554	1516790	79	SirA-like			arCOG02062	SirA	O	Predicted redox protein, regulator of disulfide bond formation	9	1
Kcr_1536	+	1516800	1517930	377	FAD-dependent pyridine nucleotide-disulphide oxidoreductase			arCOG01064	HcaD	R	NAD(FAD)-dependent dehydrogenase	10	5
Kcr_1537	+	1517941	1518360	140	protein of unknown function DUF1641			arCOG02113	-	S	Uncharacterized conserved protein	8	3
Kcr_1538	-	1518366	1518833	156	conserved hypothetical protein			arCOG02064	-	R	Peroxiredoxin family protein	8	1
Kcr_1539	+	1518933	1519733	267	translation initiation factor 2, alpha subunit			arCOG04107	SUI2	J	Translation initiation factor 2, alpha subunit (eIF-2alpha)	13	2
Kcr_1540	+	1520046	1522799	918	hypothetical protein			arCOG03418/3	0	R	S-layer-like protein	0	7
Kcr_1541	+	1522891	1525272	794	hypothetical protein			arCOG04400/1	-	M	S-layer domain	0	1
Kcr_1542	+	1525262	1525843	194	hypothetical protein								
Kcr_1543	+	1527563	1528426	288	Protein of unknown function DUF11			arCOG03270/3	0	S	Uncharacterized conserved protein, contains DUF11 domain	0	7
Kcr_1544	+	1528445	1528780	112	ribosomal protein 60S			arCOG04287/3	RPP1A	J	Ribosomal protein L12E/L44/L45/RPP1/RPP2	13	2
Kcr_1545	+	1528815	1531502	896	alanyl-tRNA synthetase	alaS		arCOG01255	AlaS	J	Alanyl-tRNA synthetase	13	2
Kcr_1546	+	1531541	1532146	202	conserved hypothetical protein			arCOG03229	-	S	Uncharacterized conserved protein	5	1
Kcr_1547	+	1532143	1533192	350	Initiation factor 2B alpha/beta/delta			arCOG01123	-	J	Translation initiation factor eIF-2B alpha subunit	13	2
Kcr_1548	-	1533164	1534162	333	Glycolate reductase			arCOG01755	LdhA	C	Lactate dehydrogenase or related 2-hydroxyacid dehydrogenase	7	7
Kcr_1549	+	1534316	1535797	494	elongation factor Tu-like			arCOG01562	GTPBP1	R	GTPase	8	7
Kcr_1550	+	1536155	1537633	493	type III restriction enzyme, res subunit			arCOG00876	SSL2	K	DNA or RNA helicase of superfamily II	6	4
Kcr_1551	+	1537674	1537898	75	FeoA			arCOG02102	FeoA	P	Fe2+ transport system protein A	3	1
Kcr_1552	+	1537895	1539940	682	ferrous iron transport protein B	feoB		arCOG00359	FeoB	P	Fe2+ transport system protein B	7	2
Kcr_1553	+	1540011	1540508	166	regulatory proteins, AsnC/Lrp			arCOG01580	Lrp	K	Transcriptional regulator (Lrp/AsnC family)	13	2
Kcr_1554	-	1540473	1541192	240	5-formyltetrahydrofolate cyclo-ligase			arCOG00474	-	H	5-formyltetrahydrofolate cyclo-ligase	13	7
Kcr_1555	+	1541232	1541549	106	transcriptional regulator consisting of an HTH domain fused to a Zn-ribbon			arCOG04479	-	K	Predicted transcriptional regulator containing an HTH domain fused to a Zn-ribbon	10	1
Kcr_1556	-	1541542	1542747	402	peptidase M20			arCOG01107	ArgE	E	Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase or related deacylase	13	2
Kcr_1558	+	1544281	1545357	359	transposase, IS605 OrfB			arCOG00679	-	L	Transposase	9	1
Kcr_1559	+	1545424	1546068	215	adenylate kinases	adk		arCOG01046	Adk	F	Adenylate kinase or related kinase	2	1
Kcr_1560	+	1546145	1546570	142	ribosomal protein L14b/L23e			arCOG04095	RplN	J	Ribosomal protein L14	13	2
Kcr_1561	+	1546580	1546975	132	ribosomal protein L24	rplX		arCOG04094	RplX	J	Ribosomal protein L24	13	2
Kcr_1562	+	1546977	1547684	236	Ribosomal protein S4E			arCOG04093	RPS4A	J	Ribosomal protein S4E	13	2
Kcr_1563	+	1547689	1548246	186	ribosomal protein L5			arCOG04092	RplE	J	Ribosomal protein L5	13	2
Kcr_1564	+	1548247	1548420	58	ribosomal protein S14			arCOG00782	RpsN	J	Ribosomal protein S14	13	2
Kcr_1565	+	1548430	1548822	131	ribosomal protein S8			arCOG04091	RpsH	J	Ribosomal protein S8	13	2
Kcr_1566	+	1548833	1549381	183	ribosomal protein L6			arCOG04090	RplF	J	Ribosomal protein L6P	13	2
Kcr_1567	+	1549392	1549739	116	Ribosomal protein L32e			arCOG00781	RPL32	J	Ribosomal protein L32E	13	2
Kcr_1568	+	1549751	1550152	134	Ribosomal protein L19e			arCOG04089	RPL19A	J	Ribosomal protein L19E	13	2
Kcr_1569	+	1550158	1550745	196	ribosomal protein L18P/L5E			arCOG04088	RplR	J	Ribosomal protein L18	13	2
Kcr_1570	+	1550770	1551402	211	ribosomal protein S5-like			arCOG04087	RpsE	J	Ribosomal protein S5	13	2
Kcr_1571	+	1551488	1551976	163	ribosomal protein L30P	L30P		arCOG04086	RpmD	J	Ribosomal protein L30	13	2
Kcr_1572	+	1551973	1552452	160	ribosomal protein L15			arCOG00779	RplO	J	Ribosomal protein L15	13	2

Gene	Strand	Start	Stop	Size (a.a.)	Product	Gene Name	COGs	arCOGs	COG_gene	Func	arCOG annotation	# of Cren in arCOGs	#
Kcr_1573	+	1552439	1553842	468	SecY protein			arCOG04169	SecY	U	Preprotein translocase subunit SecY	13	2
Kcr_1574	+	1553847	1554350	168	membrane protein-like			arCOG02673	-	S	Predicted membrane protein	1	2
Kcr_1575	+	1554465	1554764	100	Ribosomal protein L34e			arCOG04168	RPL34A	J	Ribosomal protein L34E	11	1
Kcr_1576	+	1554800	1555078	93	ribosomal protein L14E			arCOG04167	RPL14A	J	Ribosomal protein L14E/L6E/L27E	12	1
Kcr_1577	+	1555075	1556052	326	Pseudouridylate synthase			arCOG00987	TruB	J	Pseudouridine synthase	13	2
Kcr_1578	+	1556459	1556908	150	ribosomal protein S13			arCOG01722	RpsM	J	Ribosomal protein S13	13	2
Kcr_1579	+	1556918	1557418	167	ribosomal protein S4	rpsD		arCOG04239	RpsD	J	Ribosomal protein S4 or related protein	13	2
Kcr_1580	+	1557418	1557816	133	ribosomal protein S11			arCOG04240	RpsK	J	Ribosomal protein S11	13	2
Kcr_1581	+	1557882	1558820	313	hypothetical protein								
Kcr_1582	+	1558801	1559556	252	RNA polymerase, insert			arCOG04241	RpoA	K	DNA-directed RNA polymerase subunit D	13	2
Kcr_1583	-	1559705	1560853	383	Cytochrome bd-type quinol oxidase subunit 1-like			arCOG02720	CydA	C	Cytochrome bd-type quinol oxidase, subunit 1	5	1
Kcr_1584	-	1560856	1562472	539	cytochrome bd ubiquinol oxidase, subunit I			arCOG02721/3	CydA	C	Cytochrome bd-type quinol oxidase, subunit 1	5	7
Kcr_1585	+	1562730	1564031	434	major facilitator superfamily MFS_1			arCOG02682	ProP	G	permease of the major facilitator superfamily	10	5
Kcr_1586	+	1564149	1564511	121	ribosomal protein L15			arCOG00780	RPL18A	J	Ribosomal protein L18E	13	2
Kcr_1587	+	1564517	1564993	159	ribosomal protein L13	L13		arCOG04242	RplM	J	Ribosomal protein L13	13	2
Kcr_1588	+	1564977	1565396	140	ribosomal protein S9			arCOG04243	RpsI	J	Ribosomal protein S9	13	2
Kcr_1589	+	1565423	1565827	135	translation initiation factor eIF-5A	eIF_5A		arCOG04277	Efp	J	Translation elongation factor P (EF-P)/translation initiation factor 5A (eIF-5A)	13	2
Kcr_1590	+	1565832	1567145	438	GTP-binding signal recognition particle SRP54, G-domain			arCOG01228	Ffh	U	Signal recognition particle GTPase	13	2
Kcr_1591	+	1567189	1568310	374	conserved hypothetical protein			arCOG01015	-	J	Predicted pseudouridylate synthase	13	2
Kcr_1592	+	1568325	1568621	99	ribosomal protein L21e			arCOG04129	RPL21A	J	Ribosomal protein L21E	13	2
Kcr_1593	+	1568624	1568980	119	RNA polymerase Rpb4			arCOG01016	-	K	DNA-directed RNA polymerase, subunit F (rpoF)	13	2
Kcr_1594	+	1569001	1569576	192	putative nucleotide binding protein			arCOG04130	-	J	Predicted RNA-binding protein	13	2
Kcr_1595	-	1569569	1570303	245	aspartate/glutamate/uridylate kinase			arCOG00860	-	R	Predicted archaeal kinase	11	2
Kcr_1596	+	1570353	1571942	530	ATPase			arCOG04116	-	R	ATPase (PilT family)	13	2
Kcr_1597	-	1571921	1572460	180	Cytidyltransferase-related			arCOG00972	NadR	H	Nicotinamide mononucleotide adenylyltransferase	13	2
Kcr_1598	-	1572555	1573247	231	NUDIX hydrolase			arCOG01075/1	-	F	NUDIX family hydrolase	12	2
Kcr_1599	-	1573247	1574104	286	ATPase, BadF/BadG/BcrA/BcrD type			arCOG02679/2	-	I	Activator of 2-hydroxyglutaryl-CoA dehydratase (HSP70-class ATPase domain)	0	1
Kcr_1600	-	1574101	1574652	184	hypothetical protein								
Kcr_1601	-	1574669	1575298	210	metallophosphoesterase			arCOG01145	-	R	Icc family phosphoesterase	11	2
Kcr_1602	+	1575350	1576354	335	Radical SAM			arCOG00951	-	R	Predicted Fe-S oxidoreductase	9	1
Kcr_1603	+	1576412	1577896	495	phenylalanyl-tRNA synthetase, alpha subunit	pheS		arCOG00410	PheS	J	Phenylalanyl-tRNA synthetase alpha subunit	13	2
Kcr_1604	+	1577901	1579571	557	phenylalanyl-tRNA synthetase, beta subunit	pheTa		arCOG00412	PheT	J	Phenylalanyl-tRNA synthetase beta subunit	13	2
Kcr_1605	-	1579688	1580731	348	peptidase M42			arCOG01518	FrvX	G	Cellulase M or related protein	9	2
Kcr_1606	+	1580785	1582050	422	Phosphopyruvate hydratase			arCOG01169	Eno	G	Enolase	13	2
Kcr_1607	-	1582028	1583419	464	tRNA adenylyltransferase			arCOG04249	CCA1	J	tRNA nucleotidyltransferase (CCA-adding enzyme)	13	2
Kcr_1608	-	1583403	1584008	202	sugar isomerase (SIS)			arCOG00068	GutQ	M	Predicted sugar phosphate isomerase involved in capsule formation	12	1
Kcr_1609	+	1584057	1584704	216	protein of unknown function DUF125, transmembrane			arCOG01091	-	S	Uncharacterized membrane protein	10	1
Kcr_1610	+	1584689	1585387	233	tRNA pseudouridine synthase			arCOG04449	TruA	J	Pseudouridylate synthase	4	2
Kcr_1611	-	1585365	1585643	93	hypothetical protein								
Kcr_1612	-	1585742	1587238	499	Radical SAM			arCOG01355	-	C	Radical SAM superfamily enzyme	7	8
Kcr_1613	-	1587423	1588211	263	conserved hypothetical protein			arCOG04612	-	S	Uncharacterized protein conserved in archaea	0	1
Kcr_1614	-	1588208	1588738	177	Transcription factor TFIIE, alpha subunit			arCOG04270	TFA1	K	Transcription initiation factor IIE, alpha subunit	13	2
Kcr_1615	-	1589289	1589804	172	histidine triad (HIT) protein			arCOG00419	Hit	F	HIT family hydrolase	13	2
Kcr_1616	-	1589801	1590151	117	Alpha-NAC-related protein			arCOG04061	EGD2	K	Transcription factor homologous to NACalpha-BTF3	12	2
Kcr_1617	+	1590211	1590750	180	helix-turn-helix motif			arCOG01863	-	K	Predicted transcription factor, homolog of eukaryotic MBF1	12	2

Table 4. Representation of genes that are characteristic of *Crenarchaeota* or *Euryarchaeota* in. "*Candidatus* K. cryptofilum"

Locus Tag	arCOG	Functional class	Annotation	#Cren genomes	#Eury genomes
Genes characteristic of *Euryarchaeota*[a]					
Kcr_1527	**arCOG04447**	**Replication/repair**	**Archaeal DNA polymerase II, large subunit**	**0**	**27**
Kcr_1385	**arCOG04455**	**Replication/repair**	**Archaeal DNA polymerase II, small subunit**	**0**	**26**
Kcr_1284	**arCOG00872**	**Replication/repair**	**ERCC4-like helicase**	**0**	**26**
Kcr_0242	**arCOG02610**	**Replication/repair**	**Rec8/ScpA/Scc1-like protein (kleisin family)**	**0**	**24**
Kcr_1446	**arCOG02258**	**Replication/repair**	**RPA family protein, a subunit of RPA complex**	**0**	**20**
Kcr_0892, 0893, 0895, 0897, 0898, 1343, 1440	**arCOG02201**	**Cell division**	**Cell division GTPase FtsZ**	**0**	**26**
Kcr_0243	**arCOG00371**	**Cell division**	**Chromosome segregation ATPase, Smc**	**0**	**24**
Kcr_0227	arCOG01684	transcription	Predicted transcriptional regulator	0	22
Kcr_1298	arCOG03363	Energy metabolism	Archaeal/vacuolar-type H+-ATPase subunit H	0	26
Kcr_0321, 1119	arCOG01549	Energy metabolism	Coenzyme F420-reducing hydrogenase, alpha subunit	0	18
Kcr_1264	arCOG00959	Energy metabolism	Ferredoxin	0	18
Kcr_1262	arCOG01607	Energy metabolism	Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidoreductase, alpha subunit	0	19
Kcr_1100	arCOG04353	Amino acid metabolism	3-dehydroquinate synthase	0	20
Kcr_0835	arCOG02714	Coenzyme metabolism	GTP and metal dependent enzyme involved F420 coenzyme biosynthesis	0	20
Kcr_0481	arCOG00043	General functional prediction only	ATP-utilizing enzyme of the PP-loop superfamily	0	18
Kcr_1356	arCOG00354	General functional prediction only	GTPase SAR1 or related small G protein	0	23
Kcr_0479	arCOG02465	General functional prediction only	NCAIR mutase (PurE)-related protein	0	18
Kcr_0896, 1002	arCOG02155	General functional prediction only	Protein implicated in RNA metabolism, contains PRC-barrel domain	0	26
Kcr_1574	arCOG02673	Uncharacterized	Predicted membrane protein	0	27
Kcr_1412	arCOG02263	Uncharacterized	Uncharacterized conserved protein	0	24
Kcr_1416	arCOG02197	Uncharacterized	Uncharacterized conserved protein	0	23
Kcr_1046	arCOG02408	Uncharacterized	Uncharacterized conserved protein	0	20
Kcr_0480	arCOG02701	Uncharacterized	Uncharacterized conserved protein	0	18
Missing in *K. cryptofilum*: 47 euryarchaeal-specific genes					
Genes characteristic of *Crenarchaeota*[b]					
Kcr_0816	**arCOG01013**	**Translation**	**Protein containing a domain similar to ribosomal protein L13E**	**11**	**0**
Kcr_0040	**arCOG04327**	**Translation**	**Ribosomal protein S25**	**13**	**0**
Kcr_1513	**arCOG04293**	**Translation**	**Ribosomal protein S30**	**13**	**0**
Kcr_1196	**arCOG04305**	**Translation**	**Ribosomal protein S26**	**13**	**0**
Kcr_1004	**arCOG04271**	**Transcription**	**DNA-directed RNA polymerase, subunit RPB8**	**12**	**0**
Kcr_0112	**arCOG00393**	**Transcription**	**Predicted membrane-associated transcriptional regulator**	**9**	**0**
Kcr_0586	arCOG00287	Replication/repair	Predicted NurA-like nuclease	10	0
Kcr_0394	arCOG02299	Secretion/motility	Peptidase A24A, prepilin type IV	12	0
Kcr_0105	arCOG02960	General functional prediction only	Predicted aminopeptidase, Iap family	10	0
Kcr_1526	arCOG04136	General functional prediction only	Uncharacterized Zn ribbon-containing protein	12	0
Kcr_0281, 0620	arCOG01618	General functional prediction only	Aldo/keto reductase, related to diketogulonate reductase	9	0
Kcr_0852	arCOG03721	General functional prediction only	HEPN domain	10	0
Kcr_0392, 0733, 0851	arCOG03722	General functional prediction only	HEPN domain	9	0
Kcr_0387	arCOG03119	Uncharacterized	DedA family membrane protein	13	0
Kcr_1152	arCOG03770	Uncharacterized	Uncharacterized conserved protein	11	0
Missing in *K. cryptofilum*: 30 crenarchaeal-specific genes					

[a]Bold type shows genes for replication/repair and cell division system components

[b]Bold type shows genes for translation and transcription system components

Table 5. Bacterial COGs without archaeal members except "*Ca*. K. cryptofilum"

Locus Tag	COG	Function	Top BLASTP Hit	Accession #	E-value
Kcr_0172	COG0386	Glutathione peroxidase	*Desulfuromonas acetoxidans* DSM 684	ZP_01311479	7.0e-06
Kcr_0823	COG0861	Membrane protein TerC possibly involved in tellurium resistance	*Saccharopolyspora erythraea* NRRL 2338	YP_001105852	2.0e-59
Kcr_1396	COG1349	Transcriptional regulators of sugar metabolism	*Anaeromyxobacter dehalogenans* 2CP-1	ZP_02325805	6.7e-02
Kcr_0831	COG1760	L-serine deaminase	*Algoriphagus* sp. PR1	ZP_01720767	8.0e-70
Kcr_1493	COG1995	Pyridoxal phosphate biosynthesis protein	*Acidiphilium cryptum* JF-5	YP_001234273	1.0e-73
Kcr_1494	COG3395	Uncharacterized protein conserved in bacteria	*Alkaliphilus metalliredigens* QYMF	YP_001318120	5.0e-24
Kcr_1364	COG3485	Protocatechuate 3 4-dioxygenase beta subunit	*Desulfococcus oleovorans* Hxd3	YP_001530052	5.0e-06

Table 6. Top BLASTP hits for predicted euryarchaeal specific genes in "*Ca*. K. cryptofilum"

Locus Tag	Function	Top BLASTP Hit	Accession #	E-value
Kcr_0047	Histones H3 and H4	*Methanobacterium formicicum*	P48783	9.0e-14
Kcr_1137	Histones H3 and H4	*Thermococcus zilligii*	P95669	6.0e-14
Kcr_1527	Archaeal DNA polymerase II, large subunit	*Methanosaeta thermophila* PT	YP_843669	0.0
Kcr_1385	Archaeal DNA polymerase II, small subunit	*Methanosaeta thermophila* PT	YP_843089	1.0e-73
Kcr_1440	Cell division GTPase	*Thermococcus kodakarensis* KOD1	YP_184684	3.0e-68
Kcr_1343	Cell division GTPase	*Methanothermobacter thermautotrophicus* str. Delta H	NP_276787	1.0e-89
Kcr_0892	Cell division GTPase	*Thermococcus kodakarensis* KOD1	YP_184684	2.0e-38
Kcr_0893	Cell division GTPase	*Pyrococcus horikoshii* OT3	NP_142027	1.0e-28
Kcr_0895	Cell division GTPase	*Thermococcus kodakarensis* KOD1	YP_183834	2.0e-41
Kcr_0897	Cell division GTPase	*Pyrococcus abyssi* GE5	NP_126968	7.0e-29
Kcr_0898	Cell division GTPase	*Pyrococcus furiosus* DSM 3638	NP_578254	3.0e-42